UTSA Libraries
2 0000 00000010 5

V.S. Varadarajan

Reflections on Quanta, Symmetries, and Supersymmetries

Springer

WITHDRAWN
UTSA LIBRARIES

Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

https://archive.org/details/reflectionsonqua0000vara

Reflections on Quanta, Symmetries, and Supersymmetries

V. S. Varadarajan

Reflections on Quanta, Symmetries, and Supersymmetries

V. S. Varadarajan
University of California
Department of Mathematics
Los Angeles, CA 90095-1555
USA

ISBN 978-1-4419-0666-3 e-ISBN 978-1-4419-0667-0
DOI 10.1007/978-1-4419-0667-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011928909

Mathematics Subject Classification (2010): 22D10, 22E46, 34N99, 81P99

© Springer Science+Business Media, LLC 2011
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Library
University of Texas
at San Antonio

In memory of Mackey and Harish-Chandra

Contents

Preface ... ix

1 Prologue ... 1
1.1 Reality and its description ... 1
1.2 A quantum education and evolution ... 8
References ... 14

2 Quantum Algebra ... 15
2.1 The quantum algebra of Dirac ... 15
2.2 The von Neumann perspective ... 19
2.3 The measurement algebra of Schwinger ... 23
2.4 Weyl–Moyal algebra and the Moyal bracket ... 35
2.5 Quantum algebras over phase space ... 49
2.6 Moshe Flato remembered ... 51
References ... 55

3 Probability in the quantum world ... 57
3.1 The statistical interpretation of quantum theory ... 57
3.2 The uncertainty principle of Heisenberg ... 62
3.3 Hidden variables ... 65
3.4 EPR ... 74
3.5 Transition probabilities in the atom ... 79
3.6 Feynman path integrals and the Feynman–Kac formula ... 81
References ... 90

4 Super geometry ... 93
4.1 The evolution of classical geometry ... 93

4.2 Super geometry 100
4.3 The theory of super manifolds 105
4.4 Super Lie groups 116
References 118

5 Unitary representations of super Lie groups 121
5.1 Unitary representations of a super Lie group 121
5.2 Super imprimitivity theorem for even homogeneous spaces .. 129
5.3 Super semidirect products and their unitary irreducible representations 136
5.4 Super Poincaré Lie algebras and Lie groups 143
5.5 Unitary representations of super Poincaré groups 150
5.6 Super particles and their multiplet structure 152
References 153

6 Nonarchimedean physics 155
6.1 Planck scale and the Volovich hypothesis 155
6.2 Nonarchimedean symmetry groups and their multipliers 157
6.3 Elementary particles over nonarchimedean fields 165
6.4 Nonarchimedean quantum field theory 174
References 176

7 Differential equations with irregular singularities 179
7.1 Introduction 180
7.2 Reduction theory 184
7.3 Formal reduction at an irregular singularity and the theory of linear algebraic groups 187
7.4 Analytic reduction theory: Stokes phenomenon 193
7.5 The Stokes sheaf and the scheme structure on $H^1(St)$ 195
7.6 Reduction theory for families 197
References 204

8 Mackey, Harish-Chandra, and representation theory 207
8.1 George Mackey and his view of representation theory 207
8.2 Harish-Chandra as I knew him 212
8.3 Some reflections 217
8.4 Fourier analysis on semisimple groups 223
References 235

Preface

The essays in this book collect some of the lectures I have given at various places. Although the lectures ranged over a wide array of topics, they had a single theme which reflected my deep interest in the role of symmetry and supersymmetry in quantum theory. As such they are mathematical as well as personal.

Symmetry and supersymmetry, especially in quantum theory, are expressed in the language of the theory of unitary representations. This is a subject of great intrinsic beauty and enters other parts of mathematics at a very deep level. Two of the greatest figures in its history are *Mackey* and *Harish-Chandra*. Their work (to use the words of Weyl) affords shade to large parts of present day mathematics and high energy physics. It is to their memory that this volume is lovingly dedicated.

The essays should perhaps be viewed like a stroll through a garden of ideas: quantum algebras, super geometry, unitary supersymmetries, differential equations, non-archimedean physics, are a few of the topics encountered along the way. My own mathematical education evolved out of interactions with Mackey and Harish-Chandra, and I conclude this volume with brief portraits of their work, embedded in the context of personal reminiscences.

Quite a large number of people have had a hand in shaping my views expressed in these essays. In the essays themselves I have made an effort to mention some of them. But I must acknowledge my deep gratitude here to my longtime friend and collaborator Don Babbitt for the hundreds of discussions on the topics discussed here, spread literally over a lifetime.

I am also indebted to Don Blasius, who, despite many demands, personal as well as professional, on his time, took a deep interest in these essays and made many suggestions for improvement. A first draft of a substantial part

of this book was written while my wife Veda and I were vacationing in the Dordogne, living in Don's house in May of 2009. The delights of rural France were a big inspiration for me and so I should thank Don and Peter, as well as Barry and Julie (not to mention Sukie!), for making life so pleasant in Chez Blasius.

Finally, I wish to thank Ann Kostant for the help and encouragement she has given me throughout this enterprise. It was her enthusiastic acceptance of my suggestion, made in the Spring of 2009, that Springer publish a book of essays collecting together some of the lectures I have given at various places, that was the genesis of this book. I really cannot thank her enough.

Pacific Palisades
Summer 2010

Chapter 1
Prologue

The world is built with a great harmony but not always in
the form which we expect before unveiling it.
Goro Shimura

Some thoughts on reality and its description, as well as some personal recollections.

1.1 Reality and its description
1.2 A quantum education and evolution

1.1 Reality and its description

I have always admired the profound insight behind the remark of Shimura quoted above. He was actually referring to the world of mathematics[1] but I have modified his remark to mean the physical world since it makes even more sense in that context. The essays that follow, which are essentially reworkings of lectures I had given at various places, give, among other things, a highly personal view of the quantum world that has evolved in the twentieth century as an offspring of the genius of many great physicists and mathematicians. My discussions are neither complete nor entirely objective, but there are aspects to them that I feel are rather compelling. This prologue is something like an introductory discussion that sets the stage for what is to come and gives a preview of the themes to be explored. I have mixed it with some personal reminiscences of how these ideas got sorted out in my own mind.

It is generally agreed that it was *Galileo Galilei* who first insisted that the description of physical reality must be in the language of mathematics. However, the phrase *physical reality* in the above statement should not be taken in any absolute sense; I use the phrase to mean just what we are able to perceive *under current circumstances*. Thus it certainly has a provisional character, with a more comprehensive meaning to us than to Galileo. Based on

our past experience it appears that physical reality is *layered*; as our abilities to make observations improve, we peel off the existing layers to peer inside the new layer and try to come to an understanding of it. The new perceptions often force us to invent new mathematics to describe them, as well as new ways of setting up the dictionary between mathematics and phenomena, which may be completely different from the earlier ones.

I want to emphasize that this is not a question of setting up an axiomatic scheme and showing that the new theories are just natural consequences of these axioms. Indeed there is "no functor in the sky" that will reveal the secrets of the physical world to the mathematician. Rather, it is the creation of a new way of interpreting phenomena that is mirrored in the mathematical models we propose. These models seem perfect till we penetrate to the next layer, when their inadequacy is revealed, and we are confronted with new problems and new dilemmas.

Nothing illustrates this better than the transition from classical to quantum mechanics which is one of the most dramatic in the entire history of science. Many serious mathematicians and physicists believed, at the turn of the twentieth century, that the task of the natural philosopher was finished, and all that remained was the computation of the fundamental constants of nature to ever greater accuracy. But in just a short time, spectroscopic observations overwhelmed the ability of classical electromagnetic theory to explain and predict them. Indeed, according to the classical electromagnetic theory applied to the model where the electrons are circling around a nucleus, the moving electrons will radiate and continually lose energy; this would imply that the spectral lines would be continually shifting, and ultimately the electrons will fall into the nucleus. This is in stark contradiction to what is observed: *the existence of sharp spectral lines*. Thus the stability of matter could not be accounted for on the basis of classical electromagnetic theory. Clearly, deeper explanations were required. A great step was taken by Niels Bohr when he created what is now called the *Bohr atom*. For a certain period of time it was enough to work with this hybrid model where the classical structure was buttressed by a set of rules (*quantum conditions*) that took into account the new phenomena. After spectacular initial successes discrepancies began to appear and cloud the picture, as new phenomena could no longer be explained except by making artificial and *ad hoc* modifications of the theory of Bohr. Eventually it was recognized that new principles had to be introduced in describing atomic phenomena. The first step in this transition, which was indeed a gigantic one, namely the creation of a fundamentally new mechanics, was taken by Heisen-

berg. He started with the basic principle that the new mechanics should be based on only quantities that are physically observed. In the case of atomic transitions this meant to him that each physical quantity should be represented by an *infinite matrix* indexed by the energy levels E_a whose entries x_{ab} are the complex amplitudes associated with the transition from E_a to E_b. In an absolutely mysterious manner he realized that the *algebra* of matrices entered into the picture because he stipulated that if X is the matrix associated with the physical quantity x, then X^n is the matrix corresponding to the quantity $x^n (n = 1, 2, \dots)$.

The idea that matrices represented physical quantities, with its mystic origins in Heisenberg's mind, introduced a strange new complication in the mathematics because of the fact that *matrix multiplication is non commutative*. Indeed, Heisenberg believed at first that this was a weakness of his theory. But it was Dirac who understood that instead of being a weakness it was the *key to the description of the quantum world*. To Dirac the set of physical quantities, which is represented in classical mechanics by an algebra of functions, and hence a commutative algebra, is represented in quantum mechanics by a non commutative algebra, such as an algebra of matrices. Dirac called this algebra the *quantum algebra*. He made the further remarkable discovery that in the quantum algebra the commutator $ab - ba$ of two physical quantities a, b, really corresponded to the classical Poisson Bracket of the quantities interpreted classically, although there would be elements of the quantum algebra which had no classical analogs, such as spin. The structure of the quantum algebra depended on $\hbar$, Planck's constant, in such a way that when $\hbar$ could be neglected, the quantum algebra became commutative.

It must be noticed that the change in point of view from Heisenberg to Dirac, is dramatic and yet subtle. The Heisenberg matrices are very concrete and tied very intimately to very physical aspects: atomic transitions, energy levels, transition amplitudes, and so on. The Dirac algebra is very abstract, with a structure dependent on $\hbar$ that becomes commutative when $\hbar \to 0$. This extra abstraction became so characteristic of Dirac's entire approach to the description of Nature that it eventually acquired a special name: the *Dirac mode*.

If one looks back at the description of the pre quantum world, and compares it with the new ideas that were needed to describe the quantum world, it becomes clear that what had happened was a revolution in thought, in the way we describe the physical world, and in the dictionary we set up between

the abstract concepts and concrete aspects of the world. This is what I meant when I remarked earlier that we had penetrated to a new layer of reality.

The fact that in the quantum description old pictures based on spacetime such as electron orbits were thrown out made the connection between reality and the mathematics that described it very remote from intuition and experience. The consistent physical interpretation of the mathematical schemes presented deep problems, many of which were philosophical and epistemological. Bohr and Heisenberg contributed enormously to this aspect of quantum theory. With characteristic brilliance, Heisenberg realized that there are truly fundamental reasons why the orbits are unobservable: classical physics puts severe restrictions on the precision with which we can observe *both* the position and the momentum of an electron. This is because the process of measurement of the position, which typically involves a collision with a photon, introduces uncontrollable changes in the momentum of the electron. Heisenberg's analysis involved his famous thought experiment using a gamma ray microscope. He then elevated this argument into a far-reaching principle: *we cannot separate the observer from the phenomena that are being observed so that measurements of atomic systems introduce uncontrollable disturbances to them.* His famous *uncertainty principle* was a quantification of this qualitative assumption. Bohr's contribution was to emphasize the notion of *complementarity*, most clearly illustrated by the *wave-particle duality of all matter*. The difficulties of constructing a consistent theory of quantum measurement were analyzed brilliantly by von Neumann whose work revealed the *thermodynamic nature of quantum measurement*.

The state of our description of the world of elementary particles and their interactions is another illustration of what I have said, perhaps even better, since it is still unfinished and beset with problems. In the quantum mechanics of Heisenberg, Dirac, Schrödinger, Pauli, and others, there was no possibility of deriving the Bohr transition rules with (what Weyl calls) the *magic formula*

$$E_a - E_b = \hbar \nu_{ab},$$

nor could one derive the Planck radiation formula which started the quantum revolution. The reason for this is easy to understand: these are features that result from the interaction of the atom (matter) with the electromagnetic field (radiation). To derive these one has therefore to set up a scheme in which *both* the atom and the radiation field are treated quantum mechanically. This was at an entirely new level because the *radiation field is already an infinite*

dimensional system classically. Dirac was the first person to ascend to this new level, and his treatment obtained not only all these formulae but also a resolution of one of the most vexing problems of classical physics, namely the *reconciliation of the wave and particle aspects of light*. In Dirac's theory, the classical radiation field, which was a wave field, acquired particle properties on quantization. Attempts to generalize the Dirac theory into a systematic theory of quantum fields which combine quantum theory with special relativity then encountered new problems, such as divergences. Eventually many of these difficulties (but not all) were resolved. Even quantum electrodynamics, as currently understood, which is regarded as the most accurate physical theory ever constructed, is not fully acceptable to some because of the *renormalization rules* which are just recipes for hiding ugly divergences in the mathematical models that are used. People like Dirac never accepted them and one cannot entirely give up the feeling that the current status of this and other even less accurate theories is similar to the hybrid theory of the Bohr atom with its quantum rules. The standard model does summarize what is known accurately but its esthetic ungainliness appears to suggest strongly that it is provisional, and that new phenomena (perhaps coming out of the new collider at CERN) might be strong enough to suggest more radical models which are less divergent.

The string theorists try to solve all the problems by a pure intellectual effort, akin to what Dirac did when his equation predicted anti-matter, or Einstein did when he invented his theory of gravitation. Only the future can tell if string theory will join these two as a decisive way, free of singularities, of looking at elementary particles and their fields, which is as beautiful and elegant as these two theories.

Quantum theory inaugurated a striking departure in the way things are described. Before the quantum era, mathematical descriptions of physical phenomena followed ordinary experience and there was nothing mysterious about it. But quantum theory changed this in a dramatic fashion. The fact that the phenomena to be described were remote from ordinary experience forced one to invent completely new mathematical schemes that were different from anything that preceded them. In basic nonrelativistic quantum theory one needs Hilbert spaces, self-adjoint operators and their spectral resolutions, quantizations, spin structures, and so on. Later on, when quantum mechanics was combined with special relativity, the framework was enlarged to admit spacetime symmetry groups and their unitary representations. Quantum electrodynamics brought its own characteristic features: Fock space, creation and

annihilation operators, and as a culminating flourish, the concept of a *quantized field* and renormalization rules. The singularities that arise in the adaptation of these very deep mathematical theories to understand the physical world ultimately drove the physicists to enquire into the very foundations of spacetime, and change its micro-structure from that of the familiar manifolds to a supercommuting manifold, and even, a noncommuting manifold.

Some people are uncomfortable with this foray into the deepest parts of mathematics to explain Nature. They feel that the beauty of mathematics is very seductive and leads the physicist away from his true quest of describing Nature. I remember an occasion when we had invited Julian Schwinger to speak in our department. His lecture was on *the Epistemology of Modern Physics*. In a long and far-reaching discussion after the talk, Schwinger explained to us his view that the *phenomena themselves should force us to the mathematical schemes that will best explain them*. At that time he was referring to his work on the quantum measurement algebra in which he had shown, quite brilliantly, how the modern way of explaining quantum theory via vector spaces and operators is literally forced on us by the analysis of the Stern-Gerlach experiments (see the essay on quantum algebra). This is however quite opposite to the approach of the string theorists and the supersymmetricians. For them the idea is to perfect the mathematics first and *only after that* look for physical interpretations that tie it to the real world. This is the *Dirac mode* mentioned earlier (see also the essay on super geometry).

There is also some misunderstanding, mainly on the part of mathematicians, about the way mathematics is used by the physicists. It is about the role of *rigor* in physical calculations. I happen to think that rigor is not that crucial. The essential issue is whether the particular type of mathematics is the right language to use. In some sense therefore physics is more concerned with the *formal* rather than the *analytical* structure of things, more with the question whether we are using the right language, rather than in the minute details of fitting the language with reality. The situation is quite different from what it is in mathematics. Here is what Hermann Weyl says about the role of mathematics in physical sciences: *Men like Einstein or Niels Bohr grope their way in the dark toward their conceptions of general relativity or atomic structure by another type of experience and imagination than those of the mathematician, although no doubt mathematics is an essential ingredient.*

There is also a philosophical aspect to this quest that has been examined by many people such as Weyl, Schrödinger, Schwinger, and others. The creations of the mathematician reflect an esthetic that is purely internal and

yet, miraculously, these very same constructions are precisely the ones we need when we try to understand the physical world. The examples are many and well known:

(a) the theory of ordinary differential equations and their uncanny applications to the Newtonian theory of celestial motion and universal gravitation;
(b) Riemannian geometry and its use in Einstein's general theory of relativity;
(c) fiber bundles, connections, and their emergence as the basic tools in gauge theories like Yang-Mills;
(d) the use of spin geometry in Dirac's equation for the electron and in unified theories;

and so on. In several conversations I had with Harish-Chandra he often used to refer to this phenomenon as the *coincidence of the inner and outer realities*. With so many examples like these, there has emerged a point of view that *what is not forbidden must be true* and that *only beautiful theories have a chance of being also true*. Weyl, Dirac, and perhaps even Einstein, in his later years, were among the foremost believers in this principle. It is in the spirit of this philosophy that one should view the discussions in my essay on nonarchimedean physics. There it is not a question of immediately connecting p-adic mathematics with physical reality but one of building a structure that *may* have some aspects of reality in it.

In Newtonian mechanics, and even in classical electrodynamics, the phenomena studied are for the most part close to experience, and there is no difficulty in understanding the forces acting on the bodies and setting up the dynamical schemes. But in quantum mechanics and quantum field theory this is no longer true. As a consequence one is often forced to use *symmetry* as a guide in the choice of the Lagrangian or Hamiltonian to set up the dynamical equations. Consequently underlying all the discussions in this book is the massif of *representation theory*. It is the essential ingredient in all descriptions of symmetry and is present in all the essays in this book. My own understanding of physics and mathematics is built upon this theory whose power to influence mathematics and physics is unmatched. As an illustration of its power even in purely mathematical theories I discuss in one of the essays its application to the theory of differential equations.

1.2 A quantum education and evolution

My first brush with quantum theory came in 1954 when I was an undergraduate in my home town, Madras (now called Chennai, with good justification), studying in the Presidency College for a master's degree in Statistics. It was during that year that Dirac came to Madras and gave a lecture at the University. I had heard of him of course but had absolutely no idea of what he had done or why he was revered so much. I went to the talk in a big hall which was overflowing. I understood nothing, but one remark of Dirac still stands out in my mind: *the world is non commutative*. My subsequent progress in my studies had no direct connection with quantum mechanics. It was mainly probability theory at the Indian Statistical Institute in Calcutta. After my degree I went to the US, first as a post doctoral fellow to Princeton University in 1960. In the Fall of 1960 I went to the University of Washington at Seattle due to the kindness of professor Edwin Hewitt who liked some of the mathematics I did in my thesis. Seattle was a wonderful place for me. I made the acquaintances of a number of young people—Ken Ross, Albert Frodeberg, Roger Richardson, Bill Woolf, Albert Nijenhuis, to mention a few. Harish-Chandra's suggestion that I make myself thoroughly familiar with the Chevalley volumes was always ringing in my years, but they were difficult to penetrate, especially the first volume. By the greatest of good fortune, Richardson, who came from the University of Michigan, had his own handwritten notes of Hans Samelson's courses on differentiable manifolds and Riemannian geometry. They were wonderful for a beginner like me and I worked through these and finally began to understand what the global theory of Lie groups was all about. At the same time I also carefully read Dynkin's famous paper on the classification of simple Lie groups through what we now call Dynkin diagrams, as well as Weyl's proof of his famous character formula, given in his *Classical Groups* and *Group Theory and Quantum Mechanics*.

I cannot speak enough of the courtesy and accessibility of Richardson. He was already a mature mathematician, with a regular position. His beautiful work with Nijenhuis on deformations of Lie algebras was still to come. He would come almost daily to the part of the department where Woolf and I had offices and talked about mathematics. Nijenhuis and Woolf were collaborating on their extension of the Newlander–Nirenberg work to the $\mathbf{C}^{1+\alpha}$ case, and Nijenhuis would come almost daily to our little alcove. So I was in daily

contact with beautiful mathematics, and these influences were pivotal for my career.

In Seattle two things happened: the first, a minor event, was that I picked up a used copy of Dirac's book, almost new, for one dollar (I still have it); the second was that in the summer of 1961, Professor George Mackey came to Seattle as a Walker Ames Professor and delivered a series of lectures on unitary representations and quantum mechanics. I had by this time become very interested in group representations and had studied, in a desultory fashion, the famous Murray–von Neumann papers on Rings of operators and some parts of Weyl's *Classical Groups*. But the connections of representation theory with quantum theory were not there in my mind until I heard Mackey explain them. For me these lectures were a revelation, a first glimpse of the noncommutative world. I fell in love with quantum theory, and from then on, quantum theory and representation theory were to be my most beloved interests.

That summer in Seattle with its customary lovely rain-free weather, was a most idyllic time in my life. I had no duties except for teaching one summer course. Mackey lectured every day for four weeks. The lectures were a tour de force, connecting all kinds of things with a unified philosophy that was profound and inspiring. I attended the lectures, and talked with him about almost all aspects of his work. His philosopy, his perspective in both mathematics and physics, and his passion, all made the deepest impression on me. It was the single sustained personal learning experience that I have ever had in my entire scientific career.

Inspired by Mackey's lectures I hastened to read Dirac and von Neumann. From my point of view as a probabilist, I was most struck by von Neumann's wonderful conception that in quantum theory an orthocomplemented partially ordered set replaced the Boolean algebra of events of classical probability theory. This structure, whose members are the *experimental propositions*, was called by von Neumann the *quantum logic*, and, for the standard models in physics, was just the *projective geometry* of the closed subspaces of a complex Hilbert space. The concept of an observable as a self-adjoint operator arises naturally in this context and is the proper substitute for the classical notion of a random variable. But, and this was the essential difference in the quantum case, classical calculations involving more than one random variable could not be carried out in the quantum world unless the variables are *simultaneously measurable*. It was a beautiful theorem of von Neumann that quantum observables in the standard Hilbert space model are simultaneously measur-

able if and only if their operators commute with each other, or equivalently, if and only if they are all functions of a single observable. It occurred to me to ask if this theorem of von Neumann on simultaneous observability could be proved in the abstract setting of a more or less general quantum logic. I was able to do this in the summer of 1961. Then I went to the Courant Institute, where I met Warren Hirsch who was very friendly to me and we became close friends. He encouraged me to publish this work and I submitted it to the Communications in Pure and Applied Mathematics, where it appeared.

Warren and I became close friends for we shared many common interests. I was very young, far away from home, knew nobody. Warren took me under his wing and cheered me up tremendously. I will always remember his infectious laugh, sense of humor, and supreme courteousness. I shared an office with Hermann Hannish; and Warren, who was collaborating with Hermann, would come in every morning and we would chat briefly of many things. I remember his telling me that the main qualification of Byron "Whizzer" White to be a supreme court justice was his ability as a football player!! I did not know if he was joking or serious but I enjoyed it!!

Of all the people I met during my year at Courant (1962), Warren was the most vivid, the most humane, and the most interesting, to me. In retrospect I realize that he was extraordinarily generous, to have been so accessible to an unknown young mathematician (I was 25) from nowhere. He was a probabilist, but eventually branched into mathematical epidemiology, where he made fundamental contributions and became a seminal figure.

I returned to the Indian Statistical Institute at Calcutta in 1962. I have spoken elsewhere[2] about these years and the efforts of a few of us to build something lasting at the Institute. From this group only Varadhan remained that year. We ran a seminar where we discussed whatever came to our minds: quantum theory, markov processes, representation theory, and so on. Markov processes were Varadhan's great love but I was more interested in representation theory, and so we switched to that topic. As a first step(!!) we started a project of specializing Harish-Chandra's work on infinite dimensional representations to *complex* semisimple Lie algebras [2].

For me this was the start of a long period of working on Lie groups and their representations, and represented a turning point in my mathematical education and growth. The work of Harish-Chandra and the immense effort required to understand it, led me into the world of semisimple groups. No one who has not entered it can understand its incredible beauty. Here are objects which are special but are beautiful to an almost limitless degree, and

which have moved so many people to spend extraordinary amounts of time investigating them.

Why are semisimple groups so fundamental and so all-pervasive? This question seems to have no satisfactory answer. I remember vividly the Williamstown conference on representations of semisimple groups in 1972 when Harish-Chandra was giving a series of lectures on his work on the Plancherel formula for p-adic semisimple groups. The lectures were built around what he called *the Lefschetz Principle*, to the effect that *all primes must be treated on an equal footing*. He had used this principle to guide him in his work on p-adic groups by starting from his immense insight on real groups (at the infinite prime) and seeing how the theory of real groups led him to a substantial part of the harmonic analysis of the p-adic groups. He was of course aware that this approach may not give everything in the p-adic case. To buttress his own conviction perhaps, but certainly to take the audience into his confidence, he recounted a story that Chevalley had told him. The time of the story was that of the Genesis, when God and his faithful disciple, the Devil, were creating the universe. God told the Devil that he (the Devil) had a free hand in creating whatever he wanted, but there were a few things that He (God) will take care of Himself; According to Chevalley, semisimple groups were among these special things(!). Harish-Chandra, after reciting this story, added, that he hoped that the Lefschetz Principle was also among these special things!

My interest in the world of semisimple groups would last a long time, during which period the world of physics was on the back burner, so to speak. However events changed all this very abruptly. In 1983 Harish-Chandra died of a heart attack and that was an event that changed my entire intellectual landscape. The loss of someone who was a close personal friend as well as a great mentor was too much to bear and to understand. When I recovered from it eventually, I started to work on other themes in which I could create my own paradigm. The idea that I should revive my original interest in quantum theory was a natural one and an opportunity to bring it to the front arose in 1988 when I was invited to go to Genoa, Italy, to work with the physics group there, led by Enrico Beltrametti, and his younger associates Gianni Cassinelli, Piero Truini, and their students, especially Ernesto De Vito and Alberto Levrero. We called ourselves the quantum group of Genoa for fun and that visit was the first of many since then. We worked on fundamental questions of quantum theory, relativistic wave equations, extension of the

Mackey theory to orbit schemes, and so on. The interest in physics remained unabated after that.

I had, by then, also started talking about physics with Don Babbitt and among the topics we wanted to understand was the spectral theory of first order Schrödinger systems. Gradually we realized that there was no systematic treatment of the foundations of first order systems, and that there was a group theoretic approach that appeared tantalizingly new. This was the beginning of my interest in the group theoretic view of differential equations and the starting point of a very long collaboration with Don Babbitt, my longest and most sustained with anyone. It led to many long papers [3] on differential equations with *irregular singularities*, and to very happy and deep interactions with the great masters of the theory such as Yasutaka Sibuya, Bernard Malgrange, Tosihasu Kimura, Jean-Pierre Ramis, Werner Balser, and above all, Pierre Deligne. Their interest and comments [4] were responsible for some of our best work and deepest insights, although our work touched only a small part of the beautiful theory of ordinary differential equations in the complex domain. The local reduction theory and moduli of irregular systems touches in a surprising manner on semisimple groups and their orbit spaces, not only over $\mathbf{C}$ but also over *function fields*.

In UCLA I had the privilege of becoming friends with a great master, Bob Finkelstein, and learning, not only physics but also the attitudes and approaches of physicists to physical problems. The situation became even better for me in the 1990s when Sergio Ferrara, one of the world's leading authorities on super symmetry, accepted a permanent part-time position at UCLA and started to come regularly to Los Angeles. We started discussing physics and mathematics and that was how I began to get interested in Picard-Fuchs equations, regular singular differential equations, their moduli and monodromy, and most importantly, supersymmetry. Attending Sergio's lectures gave me the idea to investigate the mathematical underpinnings of unitary representations of super Lie groups, especially super Poincaré groups. This I did with the Genoa group of Gianni Cassinelli, Alessandro Toigo, and Claudio Carmeli. I must mention that the work of Kostant [5] was a great pioneering effort in super geometry. To me however, personally it was the paper of Deligne and Morgan [6] that was truly inspirational and started my own excursions into super geometry.

The p-adic story represents an entirely different direction. Of course representations of the p-adic semisimple groups have long been of interest to mathematicians. Already in the early 1960s, Gel'fand, Mautner, Bruhat,

Harish-Chandra, Jacquet, Langlands, and others saw that unitary representations of p-adic semisimple groups had deep arithmetic significance. But no one has thought much of the relevance of nonarchimedean models to physics except Beltrametti and his collaboratiors in the early 1970s. Their view was that the singularities that were the plague of quantum field theory could have arisen from the fact that the micro-structure of spacetime was radically different from what is usually assumed, for instance, that it is a real manifold. They had the idea that one should investigate the possibility that it is a p-adic manifold. This idea was revived in a big way in the late 1980s when Igor Volovich [7] proposed that in regions of spacetime whose sizes are of the order of the *Planck units*, no measurements are possible and so spacetime geometry in such regions could not satisfy the archimedean axiom. From the Volovich point of view it is a reasonable assumption that the micro geometry of spacetime is non archimedean. I came across thse ideas in 1990s when I was already aware of the work of Beltrametti and his collaborators, of Weyl's work on the commutation rules over *finite rings*. and Manin's paper on adelic physics. My interest received a big stimulus when I went to Dubna and Moscow and had the opportunity of spending time with Igor Volovich and Anatoly Kochubei. The group-theoretic connections were (and are) especially fascinating to me. Although the groups involved are *not* semisimple, the physically interesting cases of the Poincaré groups and the associated quantum field theories over p-adic fields and adelic rings present interesting challenges [8].

All of these topics are discussed in these essays. In this sense the essays must be read like a travelogue, an account of an intellectual journey. I mean this in the simplest and most naive sense: an account of some of the things that I came across which interested me, and which I hope will interest some others. The musical analogy with what I am going to say is that it is like a string quartet or quintet, a recital of some voices, which I, as a moderator, try to blend and present.

It should be clear to anyone who has read what I have to say in this prologue that I have a great indebtedness to a huge collection of friends. But above all the greatest debt I owe is to George Mackey and Harish-Chandra. In the last essay I write about their work and how my personal interactions with them allowed me a glimpse into what may be called, without any controversy, *the right way of looking at things*.

References

1. G. Shimura, *Automorphic Functions and Number Theory*, Springer Lecture Notes no., 54, 1968, p. 13 (top).
2. V. S. Varadarajan, *Some mathematical reminiscences*, Special issue dedicated to Daniel W. Stroock and Srinivasa S. R. Varadhan on the occasion of their 60th birthdays. Methods Appl. Anal. 9 (2002), no. 3, v–xviii.
3. V. S. Varadarajan, *Linear meromorphic differential equations: a modern point of view*, Bull. Amer. Math. Soc., 33(1996), 1–42.
4. P. Deligne, B. Malgrange, and J-P. Ramis, *Singularités irrégulières*, Correspondance et Documents, Soc. Math. de France, 2007.
5. B. Kostant, *Graded manifolds, graded Lie theory, and prequantization* in *Differential Geometrical Methods in Mathematical Physics*, Proc. Symp. Univ. Bonn, Bonn, 1975, Lecture Notes in Mathematics, Springer, Berlin, 1977.
6. P. Deligne, and J. W. Morgan, *Notes on supersymmetry (following Joseph Bernstein)* in *Quantum fields and Strings: a course for mathematicians*, Amer. Math. Soc., Providence, RI, 1999.
7. See[4,5] of the essay on nonaechimedean physics.
8. See[19a,19b,29b] of the essay on nonaechimedean physics.

Chapter 2
Quantum Algebra*

The ways of gods are mysterious, inscrutable and beyond
the comprehension of ordinary mortals.
Julian Schwinger

Quantum algebra was created by Dirac. Its evolution also bears the imprint of the genius of many great mathematicians and physicists such as Weyl, von Neumann, Schwinger, Moyal, Flato, and others. It has inspired developments in deformation theory, representation theory, quantum groups, and many other mathematical themes.

2.1 The quantum algebra of Dirac
2.2 The von Neumann perspective
2.3 The measurement algebra of Schwinger
2.4 Weyl–Moyal algebra and the Moyal bracket
2.5 Quantum algebras over phase space
2.6 Moshe Flato remembered

2.1 The quantum algebra of Dirac

The quotation above is Julian Schwinger's tribute to Weyl on the occasion of Weyl's birth centenary*, but it applies with even greater force to the mysterious way in which Dirac and Heisenberg slew all the dragons of classical physics and let quantum theory emerge. We can understand almost all of their thought processes but there will always be a residue of mystery to the moment of creative genius when things suddenly go to a new level of perception and imagination, and everything falls into its place as if by magic.

The term *quantum algebra* appeared for the first time in a paper of Dirac[1a, 1b] which has now become famous. Just a few months earlier Heisenberg [2]

* This essay and the next are based on lectures given at Howard University, Washington D.C., sponsored by my friend D. Sundararaman, in the 1990s.

had come up with the startling and revolutionary idea that in quantum theory physical observables must be represented by hermitian matrices which are in general of infinite order; and that if the physical observable x is represented by the matrix X, then the observable x^n is represented by the matrix $X^n (n = 1, 2, \ldots)$, the n^{th} power of X. However Heisenberg had realized that the matrices do not obey the commutative rule of multiplication that the classical observables did, and felt that this was a serious flaw in his scheme. Heisenberg had sent the proof sheets of his article to Fowler at Cambridge; and Fowler, who was at that time the thesis advisor to Dirac, passed them on to Dirac. After a study of Heisenberg's paper Dirac realized that the noncommutativity of the quantum observables, which had appeared to Heisenberg as an unwelcome aspect of the new mechanics, was in fact one of its central features, and led to a structure for the new mechanics which was a beautiful and far-reaching generalization of classical mechanics, and which had the classical mechanics as its limiting case in the correspondence limit when $\hbar \to 0$. Dirac's great conceptual insight was that the quantum observables belong to an algebra which is noncommutative but in which the *commutator*

$$xy - yx =: [x, y] \tag{1}$$

of two elements x, y, which measures the departure from their commutativity, *corresponds to*

$$i\hbar\{x, y\}$$

where

$$\{x, y\} = \sum_{1 \le r \le k} \left(\frac{\partial x}{\partial q_r} \frac{\partial y}{\partial p_r} - \frac{\partial x}{\partial p_r} \frac{\partial y}{\partial q_r} \right)$$

is the classical *Poisson Bracket*:

$$[x, y] = i\hbar\{x, y\} \tag{1a}$$

Implicit here is the assumption that the observables in both the classical and quantum theories are denoted by the same symbols but have different multiplicative structures. It is not clear from Dirac's discussion whether the commutator is *exactly equal to* the Poisson Bracket, although for the position and momentum observables

$$q_1, \ldots, q_k, p_1, \ldots, p_k$$

Dirac postulated the *exact* commutation rules

$$[p_r, q_s] = -i\hbar\delta_{rs} \qquad (1 \leq r, s \leq k) \tag{2}$$

Dirac observed that when $\hbar \to 0$, the right side goes to 0, so that the quantum algebra becomes commutative in the correspondence limit, and coincides with the classical algebra of physical observables. It will turn out eventually that the general commutator *is not equal to the Poisson Bracket* but *coincides with it up to terms of order* $\hbar^2$.

It is to be noted that the hermitian nature of the elements of the quantum algebra that represent the quantum observables is only implicit in Dirac's formulation. But the requirement (1a) which shows that the commutator of two quantum observables is not an observable but $i\hbar$ times one, is clearly compatible only with the hermitian condition.

The commutation rules (2), were also independently arrived at by Born, Jordan, and Heisenberg, and Weyl [3, 4, 5]. They give only the *kinematical* structure of the quantum algebra. To determine its *dynamical* structure Dirac proceeded as follows. The time derivatives (at a fixed time for instance) of the observables are maps

$$x \longmapsto x' = \frac{dx}{dt}$$

of the quantum algebra into itself with the following properties: item (i) $(ax + by)' = ax' + by' \quad (a, b \in \mathbf{C})$

(ii) $(xy)' = xy' + x'y$

(iii) x' is hermitian whenever x is hermitian i.e., they are *derivations* preserving the hermitian property. Dirac verified that in a *matrix algebra* every derivation preserving the hermitian property is of the form

$$x \longmapsto i[a, x] \tag{3}$$

for some hermitian element a (this requires finite dimensionality of the algebra and in that context it is actually a classical result going back to Skolem and Noether). So Dirac was led to the postulate that the dynamics of the quantum system is governed by an observable H such that

$$x' = \frac{dx}{dt} = i[x, H] \tag{4}$$

for all observables x. He naturally identified H with the observable representing the classical Hamiltonian and postulated that it is the same function of the quantum positions and momenta as the classical Hamiltonian is of the classical position and momentum coordinates. This prescription is of course ambiguous as Dirac himself knew (see §5 of [1a]); the lack of commutativity of q_r and p_r makes it unclear what should be the elements of the quantum algebra that correspond to various polynomial expressions in these variables. For the usual Hamiltonians of the form

$$\sum_{1 \le r \le k} c_r p_r^2 + V(q_1, \ldots, q_k) \qquad (c_r > 0 \text{ constants})$$

there is no ambiguity. Incidentally, this point goes back already to Heisenberg's first paper (see [1b] p. 266).

The above discussion makes it very clear that the transition from classical mechanics to quantum mechanics consists in replacing the classical *real* commutative algebra of real physical quantities by a *complex algebra with involution* $*$ (hereafter called the *quantum observable algebra* or the *quantum algebra*) whose real(=self adjoint) elements($x = x^*$) are identified with the physical quantities in the quantum theory. In his preface to the first German edition of his book [6] Hermann Weyl wrote as follows:

> ... it can justly be maintained that the essence of the new Heisenberg–Schrödinger–Dirac quantum mechanics is to be found in the fact that there is associated with each physical system a set of quantities, constituting a (complex) noncommutative algebra in the technical mathematical sense, the (real) elements of which are the physical quantities themselves (parentheses added).

It is to be noticed that in this initial formulation of Dirac there is no mention of states, Hilbert space, and so on. These would come later and would become part of the definitive framework for quantum mechanics that is due to von Neumann, in which states are identified with the unit vectors (up to a phase) of a complex Hilbert space, observables are the self adjoint operators on this Hilbert space, and the expectation value of an observable A in a state ψ is $(A\psi, \psi)$. If we restrict ourselves to *bounded* observables, then they are the self adjoint elements of the *complex* algebra of all bounded operators of the Hilbert space, thus furnishing what we might call the conventional model for the quantum algebra of Dirac. In illustrating the theory it is better to assume that the Hilbert space is even finite dimensional, as I shall do very often.

The Dirac–Heisenberg insight raised the following absolutely fundamental questions.

1. How can one make more understandable the transition from a commutative real algebra to a complex algebra with involution? Note that only the commutator in the quantum algebra has a physical meaning (of sorts); the multiplication is left hanging, so to speak.
2. How to construct quantum algebras?
3. The quantum algebra of Dirac's theory contains a quantity $\hbar$. For calculations it is identified with Planck's constant but in order to compare quantum results with classical results it is treated as a *parameter*, and the correspondence principle of Bohr is formulated as the requirement that when $\hbar \to 0$ the quantum algebra goes over to the classical algebra in the limit. Is there a way to formulate this idea mathematically?
4. To what extent is the quantum algebra uniquely determined?

In what follows we shall explore these questions and try to come up with answers that are currently accepted almost universally.

2.2 The von Neumann perspective

When quantum mechanics was discovered, the break with classical mechanics was so complete, and involved such strange and unexpected relationships between what were thought of as well understood notions like states and observables, that it was extremely difficult to accept it on any other ground except that it worked. If we accept the complex algebra with involution and identify its real elements with the quantum observables, then by general results from functional analysis the algebra can be realized as an operator algebra in a Hilbert space on which the quantum observables act as bounded self adjoint operators, so that one is very close to the usual model. But how to bridge the gap from the classical to the quantum algebra?

It was von Neumann who made the truly profound investigations that finally removed much of the mystery behind this extraordinary leap from classical to quantum mechanics. This is not the place to go into von Neumann's ideas in any detail; they are discussed with his inimitable clarity and succinctness in his wonderful book [7] which is, in my opinion, one of the great landmarks of twentieth century science. They have been discussed in many places [8, 9, 10]. Let me therefore confine myself to a brief outline.

The starting point of von Neumann is that associated to any physical system there is a structure, the set $\mathcal{L}$ of all *experimentally verifiable propositions.*

This set is partially ordered, namely, that $a \leq b$ if a implies b. There is also an involution $a \mapsto a^{\perp}$ which associates to a its negation. There are also some additional relations between these two concepts that flow directly out of the physical meaning of these concepts; I shall not consider these here because they are technical. The partially ordered set $\mathcal{L}$ was called by von Neumann the *logic of the system,*. He observed that if the system considered is *classical* then $\mathcal{L}$ is a *boolean algebra* obeying the rules of *classical logic*. In von Neumann's approach the fundamental fact that distinguishes quantum systems from the classical ones is that the logic $\mathcal{L}$ of a quantum system is *not a boolean algebra.* For technical reasons we shall work with Boolean σ-algebras, which are Boolean algebras where sums of countably many elements exist in the algebra. As long as observations are being made on a single observable or a set of simultaneously measurable observables, one operates within a single Boolean σ-algebra $\mathcal{B} \subset \mathcal{L}$. However a quantum logic $\mathcal{L}$ contains many different Boolean σ-algebras which cannot all be put inside a single Boolean σ-algebra. Examples of this are the observables that refer to the wave and particle aspects of light or electrons, more generally observables which are complementary like the position and the linear momenta. One may say that quantum logics are composed of *interlocking* Boolean σ-algebras, and the manner in which the Boolean σ-algebras combine to generate $\mathcal{L}$ contains the very essence of the *complementary* properties of the quantum system.

Clearly, in trying to understand the transition to quantum logics it will be important, even essential, to try to get a view of all possible logics and select those that are promising enough to serve as models for quantum phenomena. To do this satisfactorily one has to make additional assumptions that may or may not have good physical interpretation. Of course the additional assumptions have to be such that they will encompass many models that are useful in quantum theory. One of the simplest classes of logics is obtained if one is prepared to admit that the logic is isomorphic to a *classical projective geometry,* namely, the partially ordered set of linear subspaces of a finite dimensional vector space over some division ring. The fundamental theorem of classical projective geometry is the statement that under very general circumstances a finite dimensional logic is of this form. The vector space necessarily comes equipped with a definite sesquilinear scalar product such that $\perp$ in the logic corresponds to orthocomplementation in the geometry (this is a theorem of Birkhoff and von Neumann, see [8, 12]. If the division ring is the field $\mathbf{C}$ of complex numbers, we obtain (after some technical arguments) the conven-

tional model. To get infinite dimensional models one has to drop the finiteness condition and use infinite dimensional analogues of the classical theorem of projective geometry. There are many results and this is not the place to go into them; see however the discussions in [8, 9]. I would like to mention the following: my result on geometries of infinite dimension but with atoms which can be identified with the geometry of *finite dimensional* subspaces of an infinite dimensional vector space over a division ring; the theorem of von Neumann that a modular complemented lattice of rank ≥ 4 is isomorphic to the lattice of principal left ideals of a matrix ring over a regular ring (whose definite involutions correspond to the orthocomplementations of the lattice); as well as the theorem of Piron–Amemiya–Araki on the characterization of the standard Hilbertian logic [8, 9].

These results clearly show that the usual models of quantum theory where this vector space is a complex Hilbert space with its scalar product are not unreasonable at all. The road to treat the conventional model in all its details is now wide open; the full story is well known but technically nontrivial. In the context of Hilbert space it requires the Gleason theorem to identify the (pure) states of the system with the points of the *projective space* of the Hilbert space, so that the superposition principle of states that is emphasized by Dirac to be the corner stone of quantum theory is nothing other than the fact that the states correspond bijectively to the points of the projective space of the Hilbert space, and that a state σ is in superposition with two others $\sigma_i (i = 1, 2)$ if and only if the corresponding point p_σ is on the line in the projective space that joins p_{σ_1} and p_{σ_2}. The linear structure of the symmetries in quantum theory is a direct consequence of the classical theorem in projective geometry that all automorphisms of a projective geometry are induced by linear or semilinear transformations of the underlying vector space. This contains the theorem of Wigner that quantum symmetries are induced by unitary or antiunitary maps of the Hilbert space (see [13] where this is proved by elementary arguments without appealing to the fundamental theorems of classical projective geometry). An observable x is identified with the map $E \longmapsto x(E)$ where $x(E)$ is the experimental proposition that the value of x is in the set E, thus, a σ-homomorphism from the σ-algebra of borel subsets of the real line into the logic. In the Hilbert space context, one can use the spectral theorems of Hilbert and von Neumann to identify these in turn with self adjoint operators on the Hilbert space. In particular, the bounded observables are the self adjoint elements of the complex algebra of all bounded operators of the Hilbert

space whose involution is the adjoint operation, thus finally coming to the conventional model of the quantum algebra of Dirac–Heisenberg.

In very general mathematical terms, there are three objects: the vector space over $\mathbf{C}$ with a Hilbertian structure, the orthogonal projections in this vector space, and the algebra of operators on this vector space. Classical geometry shows that these structures are all equivalent: from the algebra one can derive the projections as generators of left ideals of the algebra. I personally believe that this is one of the the most direct routes to the mystery of the quantum algebra.

The simplest model of a quantum algebra is the algebra $\mathcal{A}$ of linear operators of a N-dimensional complex Hilbert space $\mathcal{H}$ with the adjoint † as the involution. The quantum observables are then the self adjoint operators of $\mathcal{H}$, and the values of the observables are the eigenvalues of these self adjoint operators. Every observable has at most N values, and generically, has exactly N distinct values. We can thus think of this algebra as the *quantization* of the classical algebra of functions on a set having N elements. If A is an observable, then by the spectral theorem, if a_i are its distinct eigenvalues and P_i are the orthogonal projections on the eigenspace corresponding to the a_i, the observable has the representation

$$A = \sum_i a_i P_i$$

The case $N = 2$ is especially important as it is the simplest. It arises when one is interested in studying *spin* or *polarization* experiments. The spin refers to the magnetic moment and historically goes back to the first experiments performed by Stern and Gerlach; we shall come to this aspect a little later when we discuss Schwinger's extremely original approach to the mathematical structure of the observable algebra of quantum kinematics.

In summary, the von Neumann analysis shows that under very general mathematical conditions there are essentially only two models for the set of observables of *finite* physical systems: the classical one of real functions on a set of N elements, or the quantum one of self adjoint elements of the complex algebra with involution of all operators on an N-dimensional Hilbert space, with the adjoint as the involution. It must be noted however that neither the complex (i.e., not self adjoint) elements nor their product in this complex algebra have any physical interpretation.

In my opinion it was only when Schwinger re-examined the foundations of the theory of measurement in the 1960s that a convincing physical interpretation of the full *complex* quantum algebra and its involution was discovered. It turns out that the complex algebra is actually the *algebra of quantum measurements*. Indeed, one starts by introducing *measurement symbols* for the acts of measurements. The product of the symbols corresponds to successive measurements, and the involution to the performance of the measurements in the reverse order. The symbols have certain natural relations among themselves, and so one can introduce the *measurement algebra* as the algebra generated by the measurement symbols with these relations. Under suitable restrictions this algebra is in fact the quantum algebra.

Finally, let me remark that there is the question that has always aroused great interest in the foundations of quantum theory, namely, to understand the role of the *complex* numbers. Since all physical quantities are real why should one work over the complex numbers? There is no definitive answer to this question, except that in almost all parts of the theory it is impossible not to work over $\mathbf{C}$. Things like matter–antimatter duality, charge conjugation and so on, cannot be discussed without the availability of complex conjugation. Also, even if one starts over the real numbers one can always complexify, and so real systems can always be thought of as complex systems that commute with a conjugation operator. There has been nevertheless some interest in pursuing this matter. See for instance [18], pp. 194–198, and the references therein.

2.3 The measurement algebra of Schwinger

The point of view of Schwinger is most beautifully expressed in his own words [14] (pp. 1–2):

> The classical theory of measurement is implicitly based upon the concept of an interaction between the system of interest and the measuring apparatus that can be made arbitrarily small, or at least precisely compensated, so that one can speak meaningfully of an idealized experiment that disturbs no property of the system. The classical representation of physical quantities by numbers is the identification of all properties with the results of such nondisturbing measurements. It is characteristic of atomic phenomena, however, that the interaction between system and instrument cannot be indefinitely weakened. Nor can the disturbance produced by the interaction be compensated precisely since it is only statistically predictable. Accordingly, a measurement on one property can produce unavoidable changes in the value previously assigned to another property, and it is without meaning to ascribe numerical values to all the attributes of a microscopic system. The mathe-

matical language that is appropriate to the atomic domain is found in the symbolic transcription of the laws of microscopic measurement.

The starting point of Schwinger's analysis was the famous Stern–Gerlach (SG) experiment on the magnetic moment of atoms. Let us recall that the experiment consisted in heating small pellets of silver to a very high temperature in a closed oven so that the pellets vaporize. Then the vapor is drawn out through an aperture in the oven and made into a thin beam by passing through a number of narrow slits. The high temperature of the vapor guarantees that the magnetic moments of the silver atoms in the vapor are completely randomly distributed. The atoms are then forced to pass through a magnetic field and then stopped by a screen. One would expect the atoms to be deflected differentially and in a continuous manner depending on the values of their magnetic moments, and so it would be natural to expect to get a uniform spread of deposits. Instead, what one finds are two dark spots, sharply localized, and nothing more. One is forced to the interpretation that the apparatus measures whether the magnetic moment is either "up" or "down" (these being defined as the direction of the magnetic field or its opposite), and that the disturbance caused by the apparatus has forced the atoms to split into two ensembles whose magnetic moments are either up or down.

Schwinger's idea was to build a theory of quantum kinematics from a generalization of this example. For the SG experiments the observed results are either $+1$ or -1, referring to the magnetic moment being either in the direction of the magnetic field or opposite to it. The directions of the magnetic fields are of course completely arbitrary, and one can create a measurement scheme where the atomic ensemble has all its atoms with their magnetic moments aligned one way and then pass through other measurement apparata whose magnetic fields are at arbitrary angles to this direction of magnetic moment; furthermore, one can repeat such measurements many times in succession. The statistical behavior of the atoms then lead to observable effects which completely describe the kinematics of this system.

To build the structure of the kinematics of general (finite) quantum systems Schwinger considers[14,15,16] a quantum system analogous to the spin system of SG experiments where a generic observable has N distinct values (for the actual SG case $N = 2$). Following Schwinger we consider a class of *generalized Stern–Gerlach experiments* in which one makes measurements on the system. Typically, a measurement scheme receives ensembles and measures the values of an observable A and then sorts them out according to the values of A. Various types of measurements can now be considered. If A is

an observable with distinct values $a_1, a_2, \ldots, a_N$, then one can conceive of an apparatus that allows only ensembles with the A-value a_j to enter and permits only those ensembles to emerge which have the A-value a_i; let us tentatively assign the symbol $M(a_i, a_j)$ for this measurement scheme, a special case of which is the symbol $M(a_i, a_i)$ for the measurement that permits only those with A-value a_i to enter and allows these to emerge. More generally, let A, B be two observables, in general incompatible, with values $(a_i), (b_j)$ respectively; then $M(a_i, b_j)$ is the tentative symbol for a measurement scheme that permits only the ensembles with the B-value b_j to enter and allows only those with the A-value a_i to emerge. When such measurements are carried out in succession, say $M_1, M_2, \ldots, M_r$, the composite measurement is denoted by $M_r M_{r-1} \ldots M_2 M_1$, the order of measurement being from right to left. The measurement symbols therefore form a monoid with a multiplication, and admits a natural involution † defined by

$$M(a, b)^\dagger = M(b, a), \qquad (M_r M_{r-1} \ldots M_2 M_1)^\dagger = M_1^\dagger M_2^\dagger \ldots M_r^\dagger$$

To construct the full quantum algebra one has to proceed to specify the rule for multiplying these measurement symbols. Schwinger does this and discovers that the measurement algebra, namely the algebra generated by the measurement symbols, is canonically isomorphic to the algebra of all operators in $\mathcal{H}$ in such a manner that the observables can be identified with the self adjoint elements and the involution with the adjoint.

I have said that the symbols $M(a, b)$ are only tentative. Before defining them formally let me follow Schwinger and discuss them a little more. Let us first consider the case when there is only one observable and consequently only the symbols

$$M(a_i, a_j) \qquad (1 \leq i, j \leq N)$$

From the physical interpretation it is clear that we should have

$$M(a_i, a_j) M(a_k, a_m) = \delta(j, k) M(a_i, a_m), \qquad M(a_i, a_j)^\dagger = M(a_j, a_i)$$

The relations

$$M(a_i, a_i)^2 = M(a_i, a_i), \qquad M(a_i, a_i) = M(a_i, a_i)^\dagger$$

which are special cases of the above relations are essentially a formulation of the obvious fact that if we pass an ensemble through the measurement apparatus symbolized by $M(a_i, a_i)$ twice we get the same result as by doing it

just once. Analogous interpretations lead to the more general rules involving all the $M(a_i, a_j)$. The symbols 1 and 0 represent measurement schemes that permit all ensembles to go through without any selection and schemes that permit no ensembles to enter respectively. A measurement scheme that permits ensembles only if their A-values lie in a subset E of the set $\{a_1, \ldots, a_N\}$ is denoted by

$$\sum_{i \in E} M(a_i, a_i)$$

If we take E to be the entire set of values of A we should get

$$\sum_i M(a_i, a_i) = 1.$$

Let us now consider measurements involving two observables. The sequence

$$M(a_i, a_i)M(b_j, b_j)M(a_i, a_i)$$

is particularly interesting. It is just the a_i-measurement except for the interposition of the b_j-measurement. From considerations of actual SG experiments one expects only a *positive fraction* of the ensembles to emerge from this scheme; if $p(a_i, b_j)$ is this fraction, Schwinger postulates that

$$M(a_i, a_i)M(b_j, b_j)M(a_i, a_i) = p(a_i, b_j)M(a_i, a_i)$$

and interprets $p(a_i, b_j)$ as the probability of finding an individual member of the ensemble in the state a_i when we know before the measurement that it was in state b_j. The matrix $(p(a_i, b_j))$ is *stochastic*, namely,

$$p(a_i, b_j) \geq 0, \qquad \sum_i p(a_i, b_j) = 1.$$

It is at this stage that Schwinger takes a bold step. He does not explain this in [14] or [15] but refers to it as a "leap of imagination" in [16]. He thinks of $M(a_i, a_j)$ as a 2-step process: first the entering a_j-ensemble is annihilated and then the emerging a_i-ensemble is created in its place. So he writes $M(a_i, a_j)$ as a *product* of two new symbols $\langle a_j|$ and $|a_i\rangle$:

$$M(a_i, a_j) = |a_i\rangle\langle a_j|$$

where $\langle a_j|$ refers to the annihilation of the state where A is a_j and $|a_i\rangle$ refers to the creation of the state where A is a_i. The act of measuring A then leads

to the plausible interpretations of the new symbols as follows:

$$\langle a_i|a_j\rangle = \delta(i, j), \qquad A|a_j\rangle = a_j|a_j\rangle \qquad \langle a_i|A = a_i\langle a_i|$$

It is clear now that $|a_i\rangle$ and $\langle a_j|$ have to be interpreted as *vectors*, and indeed as *dual sets of vectors*. Actually, to prepare the reader for this leap, he writes $|a, b|$ in place of $M(a, b)$ and the leap consists in splitting the symbol as the product $|\rangle\langle|$. The observable A then gets identified as

$$A = \sum_i a_i|a_i\rangle\langle a_i|$$

and we have, for any polynomial function $f(A)$, the same formula

$$f(A) = \sum_i f(a_i)|a_i\rangle\langle a_i|$$

Proceeding formally, the definition of multiplication of the symbols when two incompatible observables A and B are involved takes the form

$$M(a_i, a_j)M(b_k, b_m) = |a_i\rangle\langle a_j||b_k\rangle\langle b_m| = \langle a_j|b_k\rangle|a_i\rangle\langle b_m|$$

An actual calculation with the multiplication rule defined above shows that

$$p(a_i, b_j) = \langle a_i|b_j\rangle\langle b_j|a_i\rangle.$$

Since probabilities are to be nonnegative, it is natural to ensure this by requiring that

$$\langle a_i|b_j\rangle = \langle b_j|a_i\rangle^{\text{conj}}.$$

Notice now that $p(a, b)$ becomes *symmetric* in a and b, and hence that the matrix $(p(a_i, b_j))$ becomes *doubly stochastic*, i.e.,

$$p(a_i, b_j) = p(b_j, a_i), \qquad \sum_i p(a_i, b_j) = \sum_j p(a_i, b_j) = 1.$$

From this point on the treatment is essentially formal and Schwinger reaches his results mentioned above.

One should of course study algebras with symbols such as the $M(a, b)$. We shall do this presently. But before going in this direction it may be of interest to exhibit the measurement symbols in the conventional model. *The measurements symbols will be parametrized by the unit vectors rather than*

the states themselves. The most general measurement symbol is

$$M(\psi, \theta)$$

where ψ, θ are unit vectors in $\mathcal{H}$. It represents the measurement where only ensembles in the state defined by θ are admitted and only those in the state defined by ψ emerge. For brevity we write $[\theta]$ for the state defined by the unit vector θ. The adjoint is given by

$$M(\psi, \theta)^\dagger = M(\theta, \psi). \tag{1}$$

To take care of the redundancy of phase factors in defining the states we require

$$M(\lambda\psi, \mu\theta) = \lambda\overline{\mu}M(\psi, \theta) \qquad (|\lambda| = |\mu| = 1). \tag{2}$$

The multiplication is given by

$$M(\psi, \theta)M(\rho, \tau) = (\theta, \rho)M(\psi, \tau). \tag{3}$$

Finally if (ψ_i) is an orthonormal frame,

$$\sum_i M(\psi_i, \psi_i) = 1. \tag{4}$$

The Schwinger measurement algebra is the complex algebra $\mathcal{M}$ generated by the symbols $M(\psi, \theta)$ with the relations (2),(3) (4). It is clear that this algebra has an involutive conjugate-linear antiautomorphism satisfying (1). We write this involution as $\dagger$.

What is the structure of this algebra? We first construct a homomorphism of $\mathcal{M}$ into the algebra $L(\mathcal{H})$ of all linear operators of $\mathcal{H}$. Denote by

$$L_{\psi\otimes\theta} \qquad (\psi, \theta \in \mathcal{H}, ||\psi|| = ||\theta|| = 1)$$

the linear operator

$$L_{\psi\otimes\theta} : v \longmapsto (\theta, v)\psi.$$

We must remember here our physicists' convention regarding the scalar product $(\cdot)$. Then the homomorphism is given by

$$M(\psi, \theta) \longmapsto L_{\psi\otimes\theta}.$$

To verify that this defines a homomorphism of $\mathcal{M}$ into $L(\mathcal{H})$ one must check that the $L_{\psi\otimes\theta}$ satisfy the relations (1)–(4) where we write $L_{\psi\otimes\theta}$ in place of $M(\psi, \theta)$; this however is trivial. Thus we have a well defined homomorphism of $\mathcal{M}$ into $L(\mathcal{H})$ that takes $\dagger$ into the adjoint. It is obvious that this is surjective. The theorem of Schwinger is that it is injective also so that it is a canonical isomorphism.

To prove the injectivity it is enough to prove that $\dim(\mathcal{M}) \leq N^2$ since by the surjectivity we know that the dimension has to be $\geq N^2$. For this purpose we fix an orthonormal frame (ψ_i) in the Hilbert space $\mathcal{H}$. To the frame (ψ_i) we associate the N^2 symbols

$$M(\psi_i, \psi_j) \qquad (1 \leq i, j \leq N).$$

We now have, for any unit vectors θ, τ in $\mathcal{H}$,

$$\begin{aligned} M(\theta, \tau) &= \sum_i M(\psi_i, \psi_i) M(\theta, \tau) \sum_j M(\psi_j, \psi_j) \\ &= \sum_{i,j} (\tau, \psi_j)(\psi_i, \theta) M(\psi_i, \psi_j) \end{aligned}$$

which shows that all the $M(\theta, \tau)$ are in the linear span of the $M(\psi_i, \psi_j)$. On the other hand the rule for multiplication for the symbols, namely,

$$M(\psi_i, \psi_j) M(\psi_k, \psi_m) = \delta(j, k) M(\psi_i, \psi_m)$$

shows that this linear span is a *subalgebra* of $\mathcal{M}$. Thus

$$\mathcal{M} = \sum_{i,j} \mathbf{C} \cdot M(\psi_i, \psi_j)$$

showing that

$$\dim(\mathcal{M}) \leq N^2.$$

This finishes the proof. We have thus obtained

Theorem (Schwinger). *The measurement algebra is naturally isomorphic to $L(\mathcal{H})$ via the map*

$$M(\psi, \theta) \longmapsto L_{\psi\otimes\theta}$$

In this isomorphism $\dagger$ goes over to the adjoint.

One should note that the canonical isomorphism

$$M(\psi, \theta) \longmapsto L_{\psi \otimes \theta}$$

is essentially the "leap of imagination" of Schwinger that identifies $M(\psi, \theta)$ with $\psi \otimes \theta$ interpreted as an element of $L(\mathcal{H})$, namely $L_{\psi \otimes \theta}$. One should also note that the measurement algebra is spanned by the N^2 symbols associated to *any* orthonormal frame. We have thus completed our program of giving a full physical interpretation of the complex quantum algebra and its involution.

The above discussion does not go very much beyond what Schwinger did. Accardi carried out a much more penetrating analysis in [17]. I shall now give a brief description of some of Accardi's ideas and results.

The work of Accardi

The starting point for Accardi is a not necessarily finite dimensional $*$-algebra $\mathbf{A}$ over the reals $\mathbf{R}$ with center $\mathbf{K}$, and certain additional axioms. If $\mathbf{A}$ is such an algebra, a *projection* in $\mathbf{A}$ is an element X such that $X = X^* = X^2$. A *partition of unity* is a family (A_a) of projections such that $A_a \neq 0$ and $\sum_a A_a = 1$; such a family is *maximal* if the linear span of the A_a (which is an abelian subalgebra) is maximal in $\mathbf{A}$, namely, if $B \in \mathbf{A}$, $BA_a = A_a B$ for all a implies that $B = \sum c_a A_a$ for suitable $c_a \in \mathbf{K}$. $\mathbf{A}$ is a *Schwinger algebra* if there is an integer n and a set T of maximal partitions of unity,

$$\big(A(x)\big) = \big(A_a(x)\big)_{1 \le a \le n} \qquad (x \in T)$$

such that $\mathbf{A}$ is generated by the $A_a(x) (1 \le a \le n, x \in T)$ over $\mathbf{K}$. The integer n occurring in the definitions could be ∞. $\mathbf{A}$ is a *Heisenberg algebra* if, first of all, it is a Schwinger algebra with $|T| = 2$, and, writing the generating set as (A_a), (B_b), the $A_a B_b$ is a basis for $\mathbf{A}$ over $\mathbf{K}$. One can then show from the axioms that in a Schwinger algebra $\mathbf{A}$ we have

$$A_a(x) B_b(y) A_a(x) = p_{ab}(x, y) A_a(x) \qquad (1 \le a, b \le n, x, y \in T),$$

where $p_{ab}(x, y) \in \mathbf{K}$, $p_{ab}(x, y) \ge 0$ with

$$p_{ab}(x, y) = p_{ba}(y, x), \qquad \sum_a p_{ab}(x, y) = \sum_b p_{ab}(x, y) = 1$$

so that the matrix $P(x, y) = (p_{ab}(x, y))$ is *doubly stochastic*. It is called the *transition matrix* between $A(x)$ and $B(y)$. The additional axioms I mentioned

earlier are needed to guarantee the positivity of the $p_{ab}(x, y)$. Here positive elements of **A** are those which are sums of elements of the form A^*A.

Accardi's main concerns are the following: (a) given a doubly stochastic matrix with entries in **K**, when does there exist a Heisenberg algebra with this transition matrix? (b) given a family of transition matrices $P(x, y)_{x,y\in T}$, when does there exist a Schwinger algebra with these transition matrices? (c) How do the conventional Hilbert space models fit into this context? For (a) he finds, under suitable assumptions on **K** and the genericity condition that $p_{ab} > 0$ for all a, b, that the necessary and sufficient condition is that there is a unitary **K**-valued matrix $U = (u_{ab})$ such that $p_{ab} = u_{ab}^* u_{ab} = |u_{ab}|^2$ for all a, b. For (b) his condition (always assuming that the transition matrices are generic in the above sense) is that one should be able to choose unitary $U(x, y)$ generating the $P(x, y)$ such that for all triples $x, y, z \in T$, we have (what he calls the Schrödinger equations) $U(x, z) = U(x, y)U(y, z)$. The Hilbert space models are then constructed assuming that $\mathbf{K} = \mathbf{C}$, from these unitary matrices. The beautiful aspect of Accardi's theory is that the *probability amplitude u_{ab} appears in a natural way, and transition probabilities are the absolute squares $|u_{ab}|^2$ of these amplitudes, a philosophy that is the corner stone of physics from Dirac to Schwinger to Feynman.*

We shall now outline another approach to obtain at least some of the results of Accardi. We shall work from the very beginning with *complex* algebras equipped with an involution $*$ ($*$-algebras). The involution is conjugate linear and reverses the order of multiplication. Let then **A** be a *finite dimensional* $*$-algebra of dimension N. We denote by ℓ (resp. r) the left (resp. right) regular representation of **A**; it is faithful. Let

$$t(x) = \operatorname{Tr} \ell(x).$$

Then

$$t(xy) = t(yx)$$

and

$$\langle x, y\rangle = t(xy^*)$$

defines a hermitian form on **A**. Let **A** be a Heisenberg algebra with respect to two maximal partitions of unity, (A_a), (B_b) generating **A** over **C**. Then, following Accardi, we have the *structure constants* γ_{ab}^{cd} of the algebra **A**, defined by

$$B_b A_a = \sum_{cd} \gamma_{ab}^{cd} A_c B_d$$

where a, b, c, d etc go from 1 to n. From this formula we get

$$A_a B_b A_a = \sum_d \gamma_{ab}^{ad} A_a B_d$$

so that, since $\sum_d B_d = 1$, we have

$$\sum_d p_{ab} A_a B_d = \sum_d \gamma_{ab}^{ad} A_a B_d$$

giving

$$\gamma_{ab}^{ad} = p_{ab} \qquad (1 \leq d \leq n).$$

We shall now compute explicitly $\langle x, x \rangle = t(x^* x)$ for

$$x = \sum_{ab} x_{ab} A_a B_b.$$

Then

$$x^* x = \sum_{abcd} \bar{x}_{ab} x_{cd} B_b A_a A_c B_d = \sum_{abd} \bar{x}_{ab} x_{ad} B_b A_a B_d.$$

Hence, as $t(B_b A_a B_d) = t(A_a B_d B_b) = \delta_{bd} t(B_b A_a)$, we have

$$t(x^* x) = \sum_{abd} \bar{x}_{ab} x_{ad} t(B_b A_a B_d) = \sum_{ab} |x_{ab}|^2 t(B_b A_a).$$

To calculate $t(B_b A_a)$ we note that $\ell(B_b A_a)$ kills $A_c B_d$ if $c \neq a$, and

$$\ell(B_b A_a)(A_a B_d) = B_b A_a B_d = \sum_{uv} \gamma_{ab}^{uv} A_u B_v B_d = \sum_u \gamma_{ab}^{ud} A_u B_d.$$

Hence

$$t(B_b A_a) = \sum_d \gamma_{ab}^{ad} = n p_{ab}.$$

Thus

$$t(x^* x) = n \sum_{ab} p_{ab} |x_{ab}|^2, \qquad x = \sum_{ab} x_{ab} A_a B_b.$$

So the Schwinger-Accardi condition

$$p_{ab} \geq 0$$

is equivalent to the invariant condition that $\langle x, x \rangle$ is nonnegative definite. The genericity condition that

$$p_{ab} > 0$$

is equivalent to the invariant condition that $\langle x, x\rangle$ is *positive definite.*.

We thus define a *premeasurement algebra* as a finite dimensional $*$-algebra whose natural hermitian form is nonnegative. A $*$-*representation* of $\mathbf{A}$ is a representation ρ of $\mathbf{A}$ in a finite dimensional Hilbert space such that $\rho(x^*) = \rho(x)^*$ for all $x \in \mathbf{A}$. A *measurement algebra* $\mathbf{A}$ is a premeasurement algebra whose natural hermitian form is strictly positive definite. Our aim now is to show that measurement algebras are nothing more than direct sums of full matrix algebras. The results follow simply from the classical theory of finite dimensional algebras (Wedderburn) and so we will be brief.

Suppose that $\mathbf{A}$ is a measurement algebra. From $t(xy^*) = t(y^*x)$ we get $\langle x, y\rangle = \langle y^*, x^*\rangle$, and from $t(xyz) = t(yzx)$ we get

$$\langle ax, y\rangle = \langle x, a^*y\rangle, \quad \langle xa, y\rangle = \langle x, ya^*\rangle.$$

In other words, if we view $\mathbf{A}$ as a Hilbert space with scalar product $\langle x, y\rangle$, both the left and the right regular representations are $*$-representations. If M is an ideal (resp. left, right, two-sided), so in $M^\perp$. Moreover, the action of $\mathbf{A}$ on M (resp. left, right, or two-sided) is a $*$-representation. So we can write $\mathbf{A}$ as a direct sum of minimal two-sided ideals M_i.

Let M be a two-sided ideal of $\mathbf{A}$. Then $M^\perp$ is also a two-sided ideal and $\mathbf{A} = M \oplus M^\perp$. Let P be the orthogonal projection from $\mathbf{A}$ to M and let $P1 = e \in M$. Since P commutes with left and right multiplications, we have $P = \ell(e) = r(e)$; moreover, as $P = P^* = P^2$ we see that e is in the center of $\mathbf{A}$ and $e = e^* = e^2$. If now $x \in M$, we have $x = ex$ so that, taking adjoints, $x^* = x^*e$, showing that $x^* \in M$. Hence M is a $*$-ideal, i.e., it is closed under $*$. In particular we may regard M as a nondegenerate $*$-algebra with e as its unit element.

Consider the orthogonal decomposition $\mathbf{A} = \oplus_i M_i$ into minimal two-sided ideals. Let us fix i and let us take a *minimal* left ideal $N \subset M_i$. Then ℓ acts on N as an irreducible $*$-representation. Hence, ℓ is a $*$-homomorphism of $\mathbf{A}$ onto the $*$-algebra $\mathrm{End}(N)$. Now for $x \in M_i$, $y \in M_i^\perp$, xy lies in $M_i \cap M_i^\perp$ so that $xy = 0$. Hence $M_i^\perp$ is contained in the kernel K of ℓ acting on N. But $K \neq \mathbf{A}$ and so, as M_i is minimal, $K \cap M_i = 0$, showing that $K = M_i^\perp$. We may thus regard the left regular representation of M_i acting on N as a $*$-isomorphism of M_i with $\mathrm{End}(N)$. We thus see that a measurement algebra is a direct sum of matrix $*$-algebras. The converse is trivial. If the algebra is *simple*, or if it is central, i.e., its center is $\mathbf{C}$, then it is a full matrix $*$-algebra.

Since it is obvious that that direct sums of full matrix algebras are Schwinger algebras, we have justified our definition of measurement algebras.

The generic Heisenberg algebras of Accardi are simple measurement algebras. If M is a nonzero two-sided ideal and contains the element $x = \sum_{ab} x_{ab} A_a B_b$ where $x_{cd} \neq 0$, then $A_c x B_d = x_{cd} A_c B_d \in M$ so that $A_c B_d \in M$. Hence $A_c B_d A_c = p_{cd} A_c \in M$, showing that $A_c \in M$, hence $B_b A_c B_b \in M$, hence finally that $B_b \in M$. Summing over b we get $1 \in M$. Hence for these $*$-algebras we have a *Hilbert space model* for $\mathbf{A}$.

Suppose $\mathbf{A}$ is a full matrix $*$-algebra. If (A_a) is a maximal partition of unity, we have an ON basis of M, say (e_a) such that A_a are the orthogonal projection $M \longrightarrow \mathbf{C}e_a$. If we have a family $(A_a(x))_{x \in T}$, we have a family of ON bases $(e_a(x))_{x \in T}$ and it is clear that $p_{ab}(x, y) = |u_{ab}(x, y)|^2$ where $U(x, y) = (u_{ab}(x, y))$ is unitary with $u_{ab}(x, y) = \langle e_a(x), e_b(y) \rangle$. Then

$$U(x, x) = I, U(x, y)U(y, z) = U(x, z) \qquad (x, y, z \in T).$$

Conversely, suppose we are given a family $(U(x, y))_{x,y \in T}$ of unitary matrices satisfying these relations. We select an n-dimensional Hilbert space M and an ON basis $(e_a(x_0)$ of it for a fixed $x_0 \in T$ and define

$$e_a(x) = U(x_0, x)^{-1} e_a(x_0).$$

As $U(x, y) = U(x_0, x)^{-1} U(x_0, y)$ we find that $U(x, y)e_a(y) = e_a(x)$ and $p_{ab}(x, y) = |\langle e_a(x), e_b(y) \rangle|^2$. This completes the treatment of Accardi's theory in the finite dimensional nondegenerate case.

If $\mathbf{A}$ is only a premeasurement algebra, $\langle x, x \rangle$ is only nonnegative. The null vectors form a two-sided ideal $\mathbf{N}$ and $\mathbf{A}/\mathbf{N}$ is a measurement algebra. The above analysis applies to it.

It may be of interest to extend this analysis to the infinite dimensional context that Accardi considers and compare his treatment to the limits of finite dimensional measurement algebras that Schwinger treats in his papers[14].

Let us now conclude by discussing briefly a class of finite models. They are constructed out of finite abelian groups and go back to Weyl and Schwinger[6,19]. Let G be a finite abelian group and let $\widehat{G}$ be its dual group. The Hilbert space of the quantum system is $\mathcal{H} = L^2(G)$, the space of all complex functions on G with the scalar product

$$i(f, g) = \frac{1}{|G|} \sum_{x \in G} f(x) \overline{g(x)}$$

Then there are two sets of orthonormal bases for $\mathcal{H}$, the basis of delta functions $(|G|^{1/2}\delta_x)_{x\in G}$ and $(\chi)_{\chi\in\widehat{G}}$. Let A and B be two observables with distinct sets of values $(a_x)_{x\in G}$ and $(b_\chi)_{\chi\in\widehat{G}}$ and corresponding eigenvectors as the two bases. Then

$$|(|G|^{1/2}\delta_x, \chi)|^2 = \frac{1}{N}$$

so that

$$M(a_x)M(b_\chi)M(a_x) = \frac{1}{N}M(a_x)$$

In other words, in a state in which A has a sharp value, when we make a measurement of B and then see what is the A-value is after the B-measurement, we find that all the A-values are equally likely. Thus all information about the first A-measurement has been erased by the B-measurement. The relationship between A and B is symmetrical and it is clear why we should regard them as complementary to each other. The simplest case is when $G = \widehat{G} = \mathbf{Z}_N = \mathbf{Z}/N\mathbf{Z}$, the additive group of integers mod N; this was the case studied by Weyl and Schwinger. It is obvious that we have a nondegenerate Heisenberg algebra here.

2.4 Weyl–Moyal algebra and the Moyal bracket

We now turn to the remaining questions raised earlier, namely, the construction of the quantum algebra for quantum systems of finitely many degrees of freedom, its uniqueness, and its dependence on $\hbar$. The key to this investigation is the *Weyl quantization map*. Its construction goes back to his great 1927 paper, explicated later in his famous book [6]. Weyl was unable to prove the uniqueness of the quantum algebra given the relations (2) and this was done by von Neumann and Stone independently [22, 23]. I shall refer to their works in more detail later on.

Weyl's projective representation and its uniqueness

Before I start explaining the work of Weyl which is a great landmark in the subject, it is of interest to quote from Schwinger's recapitulation of his own work on measurement algebras in [16]:

> . . . Very early in my study of physics, Weyl became one of my gods. I use the word "god" rather than, say, "outstanding teacher" for the ways of gods are mysterious, inscrutable, and beyond the comprehension of ordinary mortals. . .

Already, even at the very inception of quantum theory [5], Weyl had made the suggestion that the Heisenberg–Dirac commutation rules (2) be replaced by the commutation rules between the corresponding *unitary operator groups*

$$\left(e^{iaq_l}\right), \qquad \left(e^{ibp_k}\right)$$

Let

$$U(a) = e^{ia\cdot q}, \qquad V(b) = e^{ib\cdot p} \qquad (a, b \in \mathbf{R}^k) \tag{5}$$

where

$$a\cdot q = a_1 q_1 + \cdots + a_k q_k, \qquad b\cdot p = b_1 p_1 + \cdots + b_k p_k$$

Then the relations (2) are formally equivalent to

$$U(a)V(b) = e^{-i\hbar a\cdot b} V(b)U(a) \tag{6}$$

The commutation rules (6) have the advantage of dealing with unitary operators which are better behaved than the *unbounded* operators that must represent p and q.

The essential originality of Weyl consisted in considering not just U and V by themselves but to *treat them both together.* To this end Weyl started with the map

$$(a, b) \longmapsto U(a)V(b)$$

Although the $U(a)$ and $V(b)$ are unitary representations of $\mathbf{R}^k$, the fact that U and V do not commute shows that this map is *not* a unitary representation of $\mathbf{R}^{2k}$. But it is a *projective unitary representation*, i.e., a unitary representation up to phase factors, or equivalently, a *homomorphism* of $\mathbf{R}^{2k}$ into the projective unitary group of the Hilbert space of the system. This is quite appropriate since it is the structure of this projective space with its probability functional $|(\psi, \theta)|^2$ that determines the structure of quantum mechanics.

Let me recall some basic facts about projective (unitary) representations. Given a separable locally compact topological group G, a projective representation of G is a map

$$L : x \longmapsto L_x$$

of G into the unitary operators of a Hilbert space such that the induced map into the projective unitary group is a continuous homomorphism. In down-to-earth language this is the same as saying that

$$L_x L_y = \lambda(x, y) L_{xy} \qquad (x, y \in G)$$

where

$$\lambda(x, y) \in T$$

T being the multiplicative group of complex numbers of absolute value 1. The continuity assumptions are equivalent to requiring that $x \longmapsto (L_x\psi, \theta)$ is Borel for all ψ, θ in the Hilbert space. The function $\lambda : G \times G \longrightarrow T$ is called the *multiplier* of the projective representation. The map of G into the projective unitary group of the Hilbert space is unchanged if we replace L_x by $a(x)L_x$ for any borel map $A : G \longrightarrow T$; then λ changes to $x, y \longmapsto \lambda(x, y)a(x)a(y)/a(xy)$.

Following von Neumann[23] we consider the normalized map

$$W : (a, b) \longmapsto e^{(i\hbar/2)a\cdot b} U(a)V(b) \tag{7}$$

which is also a projective representation of $\mathbf{R}^{2k}$. This normalization is also implicit in Weyl's 1927 paper (see formula immediately before (52) of that paper). A simple calculation shows that

$$W(a, b)W(a', b') = m(a, b : a', b')W(a + a', b + b') \tag{8}$$

where, m, the so-called *multiplier* of the projective unitary representation W, is given by

$$m(a, b : a', b') = e^{(i\hbar/2)(a'\cdot b - a\cdot b')} \tag{9}$$

Weyl postulated that *quantum kinematics is described completely by the projective unitary representation* W. It is clear from this that the infinitesimal generators Q_r, P_r of the unitary groups

$$U(0, \ldots, t, \ldots), \quad V(0, \ldots, t, \ldots) \qquad (t \text{ is in the } r^{\text{th}} \text{ place})$$

satisfy the commutation rules (2). He also showed that if one takes *any projective unitary representation* of a real vector space V of finite dimension d and suppose that it is *faithful and irreducible*, then its multiplier m is

necessarily of the form

$$m(x : y) = e^{(i\hbar/2)\beta(x:y)} \qquad (x, y \in V)$$

where β is a *symplectic form* on $V \times V$. It then follows that $d = 2k$ is even and one can choose a linear isomorphism of V with $\mathbf{R}^k \times \mathbf{R}^k$ such that (identifying V with $\mathbf{R}^k \times \mathbf{R}^k$)

$$\beta(a, b : a', b') = a' \cdot b - a \cdot b'$$

This means that one can restrict attention to projective unitary representations W with multiplier (9).

This was Weyl's formulation of *matrix mechanics*. A special choice of W can now be introduced, the so-called *Schrödinger model*. Here the Hilbert space $\mathcal{H}$ is $L^2(\mathbf{R}^k)$ and $W = W_0$ is defined by

$$(10) \qquad (W_0(a, b)\psi)(q) = e^{(i\hbar/2)a \cdot b} e^{ia \cdot q} \psi(q + \hbar b) \qquad (q \in \mathbf{R}^k)$$

Weyl then formulated *wave mechanics* as the kinematics determined by the representation W_0 and raised the problem of equivalence of wave and matrix mechanics in the following form: is every irreducible projective unitary representation of $\mathbf{R}^k \times \mathbf{R}^k$ with multiplier m as in (9) unitarily equivalent to W_0? It is of course quite easy to check that W_0 is irreducible.

In this form Weyl was not able to prove the uniqueness. But he made a remarkable observation, namely, that if one replaces the space $\mathbf{R}^k \times \mathbf{R}^k$ by the *finite group* $\mathbf{Z}_k \times \mathbf{Z}_k$ where $\mathbf{Z}_k$ is the additive group $\mathbf{Z}/k\mathbf{Z}$ of intergers mod k, and m by

$$m_k(a, b : a', b') = e^{2\pi i(a' \cdot b - a \cdot b')/k}$$

then the same problem of uniqueness can be formulated in this context. He then showed in [6] that the answer to the uniqueness question in the finite context is affirmative. He also suggested that in the limit as $k \to \infty$ it is possible to view the Schrödinger model as the limit of these finite quantum systems, so that the uniqueness in the general case is very plausible.

Direct proofs of the uniqueness of Weyl kinematics were constructed by Stone [22] and Von Neumann [23] very soon after Weyl's paper [6] was published. They established uniqueness in the following strong form: *any projective unitary representation of* $\boldsymbol{R}^k \times \boldsymbol{R}^k$ *with multiplier* m *defined by*

(9) *is unitarily equivalent to the direct sum of a number of copies of the Schrödinger model* W_0. Subsequently Mackey formulated and proved the uniqueness theorem in the context of any separable locally compact abelian group, and this theorem had far-reaching consequences [24].

Weyl quantization map

The concept of a quantization map is of course fundamental because the quantum Hamiltonian is obtained by applying this map to the classical Hamiltonian. Preferring to work with bounded operators, Weyl defined the quantization map $Q_\hbar$ on a space of sufficiently nice functions on $\mathbf{R}^k \times \mathbf{R}^k$. He required that

$$Q_\hbar : e^{ia\cdot q} e^{ib\cdot p} \longmapsto W(a, b) = e^{(i\hbar/2)a\cdot b} U(a)V(b)$$

and then extended it by linearity. Using Fourier analysis this means that $Q_\hbar$ is well defined for all functions on the algebra $\mathcal{F}$ which are *Fourier transforms of integrable functions of the dual variables* a, b. More precisely, let $f = \mathbf{F}g$ where $\mathbf{F}$ is the Fourier transform map:

$$f(q, p) = (\mathbf{F}g)(q, p) = \int e^{i(a\cdot q + b\cdot p)} g(a, b) da db.$$

Then $Q_\hbar(f) = W(g)$:

$$\begin{aligned} f &= \mathbf{F}g \\ Q_\hbar(f) &= \int g(a, b) W(a, b) da db. \end{aligned} \tag{11}$$

Here da etc refers to the *self-dual* Lebesgue measure on $\mathbf{R}^k$ which is $(2\pi)^{-k/2}$ times the standard Lebesgue measure.

The Weyl quantization map is thus

$$Q_\hbar : f \longmapsto W(\mathbf{F}^{-1} f) \qquad (f \in \mathcal{F}) \tag{12}$$

Weyl actually did not insist that the Fourier integrals be convergent but be defined only in some distributional sense. This allowed him to calculate formally the quantization of *polynomial functions* in q, p. He found that the quantization of a polynomial is obtained by replacing q_r, p_r by the corresponding quantum mechanical operators Q_r, P_r, and replacing any monomial in the q_r, p_r by the *averages of the corresponding products in all possible*

orderings (*the Weyl ordering*)[23]:

$$Q_\hbar(f) =: f(q, p) :$$

Furthermore, it is easy to see that

$$W(a, b)^\dagger = W(a, b)^{-1} = W(-a, -b)$$

so that

$$\int g(a, b)W(a, b)dadb \text{ is self adjoint } \Longleftrightarrow g(a, b) = g^{\text{conj}}(-a, -b)$$

which leads to

$$Q_\hbar(f) \text{ is self adjoint } \Longleftrightarrow f \text{ is real}$$

If we make additional restrictions on f, for instance that f be smooth and rapidly decreasing, i.e., that f is in the *Schwartz space*, then $Q_\hbar(f)$ is an integral operator of trace class with kernel

$$K_\hbar(f)(q, q') = \widehat{g_1}\Big((\hbar/2)(q + q'), q' - q\Big) \tag{13}$$

where $\widehat{g_1}$ is the partial Fourier transform of $g = \mathbf{F}^{-1} f$ in the first variable.

Going back to W let me recall that if A is a locally compact abelian group and w is a unitary representation of A, then the map

$$g \longmapsto w(g) := \int w(x)g(x)dx$$

is a $*$-homomorphism of the *convolution algebra* $L^1(A)$ into the algebra of bounded operators in the Hilbert space of w with the property that

$$w(g^*) = w(g)^\dagger \qquad g^*(x) = g(-x)^{\text{conj}}$$

If now w is only a projective unitary representation with multiplier δ, then Weyl observed that

$$g \longmapsto w(g) := \int g(x)dx$$

is still a $*$-homomorphism on $L^1(A)_\delta$ where $L^1(A)_\delta$ is the *twisted convolution algebra* associated to δ, i.e., the underlying vector space of $L^1(A)_\delta$ is $L^1(A)$ with the twisted convolution defined by

$$(14)\qquad \begin{aligned}(g_1 *_\delta g_2)(x) &= \int g_1(y)g_2(x-y)\delta(y, x-y)dy \\ &= \int g_1(x-y)g_2(y)\delta(x-y, y)dy,\end{aligned}$$

while the $*$-property corresponds to

$$w(g^*) = w(g)^\dagger \qquad g^*(x) = g(-x)^{\text{conj}}\delta(x, -x)$$

Let us now specialize this to the case when

$$A = \mathbf{R}^k \times \mathbf{R}^k, \delta = m$$

and write

$$g_1 *_m g_2 = g_1 *_\hbar g_2$$

which is justifiable since m is determined uniquely by $\hbar$. Then

$$(15)\quad (g_1 *_\hbar g_2)(a, b) = \int g_1(a', b')g_2(a-a', b-b')e^{(i\hbar/2)(a\cdot b' - a'\cdot b)}da'db'$$

Thus

$$(16)\qquad Q(\mathbf{F}g_1)Q(\mathbf{F}g_2) = Q(\mathbf{F}g), \qquad g = g_1 *_\hbar g_2$$

or

$$Q(f_1)Q(f_2) = Q(f)$$

where

$$f(q, p) = \int \widehat{f_1}(a', b')\widehat{f_2}(a-a', b-b')e^{(i\hbar/2)(a\cdot b' - a'\cdot b) - i(a\cdot q + b\cdot p)}da'db'dadb$$

with

$$\widehat{g}(a, b) = \int g(q, p)e^{i(a\cdot q + b\cdot p)}dqdp$$

Weyl quantization on an abelian group

The completely group theoretic nature of his quantization of the classical phase space, which is so obvious in the above discussion, led Weyl to a far-reaching generalization of the process of quantization. He formulated in his book [6] (pp. 275–276) the general principle of quantization as follows.

> The kinematical structure of a physical system is expressed by an irreducible Abelian group of unitary ray rotations in system space. The real elements of the algebra of this group are the physical quantities of the system; the representation of the abstract group by rotations of system space associates with each such quantity a definite hermitian form which "represents" it.

Here "group of ray rotations" is Weyl's term for a subgroup of the projective unitary group and so is the same as a projective unitary representation. Replacing the classical phase space $\mathbf{R}^{2k}$ by an abelian group A we can formulate Weyl's principle of quantization as follows. Let $\mathcal{F}$ be an algebra of complex-valued functions on a separable locally compact abelian group A. One need not be very specific as to how $\mathcal{F}$ is to be restricted but as in most of Weyl's examples we take $\mathcal{F}$ to a subalgebra of the algebra of functions on A which are the *Fourier transforms* of integrable functions on the dual group $\widehat{A}$. Then *quantization is a linear and faithful map*

$$Q_\hbar : f \longmapsto Q_\hbar(f) \qquad (f \in \mathcal{F})$$

from $\mathcal{F}$ into the algebra of bounded operators on a Hilbert space $\mathcal{H}$, the Hilbert space of the quantum system. If $\mathcal{F}_\hbar$ is the image of $\mathcal{F}$ by $Q_\hbar$, he viewed $\mathcal{F}_\hbar$ as the quantization of $\mathcal{F}$. The map $Q_\hbar$ was required to satisfy the *self adjoint property,* namely,

$$Q_\hbar(f^{\text{conj}}) = Q_\hbar(f)^\dagger \qquad (f \in \mathcal{F})$$

This means that $\mathcal{F}_\hbar$ has an involution, namely the adjoint operation †, and that the quantum observable that corresponds to a real classical observable f is a self adjoint element of $\mathcal{F}_\hbar$.

Although Weyl treated only the cases of a vector space or a finite group, one can clearly work with *any* separable locally compact abelian group. This was done in my paper [25]. Let me now sketch briefly the main results.

For any separable locally compact abelian group G let $Z^2(G)$ be the commutative group of Borel 2-cocycles with values in the circle group, $B^2(G)$ for the subgroup of trivial cocycles, and $H^2(G) = Z^2(G)/B^2(G)$ for the quotient group of multipliers. For any $m \in Z^2(G)$ we can give $L^1(G)$ the structure of a Banach algebra by defining the *twisted convolution* by

$$(f_1 *_m f_2)(x) = \int_A f_1(y) f_2(x - y) m(y, x - y) dy \qquad (x \in G)$$

This algebra structure depends only on the cohomology class $[m]$ of m, namely on the image of m in $H^2(G)$; and is commutative if and only if m is trivial, in which case it is isomorphic to the usual convolution algebra structure on $L^1(G)$. Let $C^2(G)$ be the group of *bicharacters* of G, $S^2(G)$ the subgroup of symmetric ones and $\Lambda^2(G)$ the subgroup of alternating ones ($m(x, x) = 1, m(x, y)m(y, x) = 1$). One has $C^2(G) \subset Z^2(G)$ and if G has no 2-torsion (this is equivalent to the dual group $\widehat{G}$ having no 2-torsion) then $C^2(G)$ maps onto $H^2(G)$ with kernel $S^2(G)$ while $C^2(G) = S^2(G) \oplus \Lambda^2(G)$ so that under the natural map of $\Lambda^2(G) \longrightarrow Z^2(G)$ we have

$$\Lambda^2(G) \simeq H^2(G) \qquad (G \text{ has no 2-torsion}).$$

Let now A be a separable locally compact abelian group with character group $\widehat{A}$. Let $\mathcal{W}(A)$ be the space of all functions on A which are in $L^1(A)$ and whose Fourier transforms are in $L^1(\widehat{A})$. (It is possible to work with the larger space of functions which are Fourier transforms of elements of $L^1(\widehat{A})$ but $\mathcal{W}(A)$ is preferable as it is mapped isomorphically onto $\mathcal{W}(\widehat{A})$ by the Fourier transform, so that it is self dual.) It is easy to see that $\mathcal{W}(A)$ and $\mathcal{W}(\widehat{A})$ are both closed under convolution, multiplication, translations by group elements, and multiplications by characters; moreover, if h, k are continuous functions on A (resp. $\widehat{A}$), then $h * k$ is in $\mathcal{W}(A)$ (resp. $\mathcal{W}(\widehat{A})$). Then, for any $\widehat{m} \in Z^2(\widehat{A})$ $\mathcal{W}(A)$ has the structure of an associative algebra given by

$$f_1 \cdot_{\widehat{m}} f_2 = g, \qquad \widehat{g} = \widehat{f}_1 *_{\widehat{m}} \widehat{f}_2$$

The structure of this algebra depends only on the cohomology class of $\widehat{m}$, and is abelian if and only if $\widehat{m}$ is trivial, in which case it is just ordinary multiplication. If A has no 2-torsion, we can take $\widehat{m} \in \Lambda^2(\widehat{A})$, the $*$-operation is $f \mapsto f^{\text{conj}}$ on $\mathcal{W}(A)$, and the trivial case corresponds to $\widehat{m} = 1$. The *generic case* is when $\widehat{m}$ is *symplectic*: this means that (in the general case when $\widehat{m}$ is not necessarily a bicharacter) that

$$\frac{\widehat{m}(x, y)}{\widehat{m}(y, x)},$$

which is always a bicharacter, induces an isomorphism of A with $\widehat{A}$ (when A has no 2-torsion and $\widehat{m}$ is in $\Lambda^2(\widehat{A})$, this is the case when $\widehat{m}$ itself induces an isomorphism of A with $\widehat{A}$). In this case it is appropriate to call $\mathcal{W}(A)$ the *Weyl algebra* of A.

We have already seen that when $A = \mathbf{R}^{2k}$ the symplectic multiplier leads to the usual Weyl–Heisenberg representations. Weyl also discussed the finite case in [6]. However recent work in physics suggests that it may be of interest to consider situations where the real manifolds are replaced by p-adic ones which arise in number theory. This leads to what I call the *arithmetic Weyl algebras.* An example of these is as follows [25]. Let K be a *nonarchimedean local field* of characteristic $\neq 2$, let V be a vector space of finite *even* dimension over K, β a symplectic bilinear form on $V \times V$, and let χ be a nontrivial additive character on K which we use to define the Fourier transform on the space $\mathcal{S}(V)$ of Schwartz–Bruhat functions on V. Then $\mathcal{S}(V)$ is the arithmetic Weyl algebra with multiplication defined by

$$(f_1 \cdot_\chi f_2)(x) = \int \widehat{f_1}(\eta)\widehat{f_2}(\xi - \eta)\chi(\beta(\eta - x, \xi))d\eta d\xi$$

The theory of quantum systems over local and global fields is very interesting but this is not the place to go into them.

Weyl–Moyal algebra and the Moyal bracket

Weyl's ideas were taken up more than twenty years later in 1949 by Moyal. In a remarkable paper [26] Moyal made a deep study of Weyl's quantization map. Moyal's motivation was to reintroduce the classical phase space in quantum mechanics by identifying the quantum algebra with an algebra of functions of the "non commuting quantities q_r, p_s". He did this by pulling back the operator product via the Weyl quantization map to a product operation on the space of functions on the classical phase space. The resulting algebra is the *Weyl-Moyal algebra.* This is of course nothing other than the Fourier transform of Weyl's twisted convolution and so we call it the *Weyl–Moyal product. Unlike Weyl who did not investigate the Poisson structure on the classical function space when it is quantized, Moyal went further and calculated the operator commutator in terms of this product and obtained what we now call the Moyal bracket on the classical function space* $\mathcal{W}(\mathbf{R}^k \times \mathbf{R}^k)$.

The Weyl–Moyal product

$$f = f_1 \cdot_\hbar f_2$$

is defined by first convolving the Fourier transforms in the twisted manner and then applying the inverse Fourier transform:

(17)

$$f(q, p) = \int \widehat{f_1}(a', b')\widehat{f_2}(a-a', b-b')e^{(i\hbar/2)(a\cdot b'-a'\cdot b)-i(a\cdot q+b\cdot p)}da'db'dadb.$$

Then

(18)

$$Q_\hbar(f_1 \cdot_\hbar f_2) = Q_\hbar(f_1)Q_\hbar(f_2), \qquad Q_\hbar(f_1 \cdot_\hbar f_2 - f_2 \cdot_\hbar f_1) = [Q_\hbar(f_1), Q_\hbar(f_2)].$$

Since the map $f \longmapsto Q_\hbar(f)$ is injective, it follows from (17), (18), and the associativity of operator multiplication that

(19) $$f_1, f_2 \longmapsto f_1 \cdot_\hbar f_2, \qquad f_1, f_2 \longmapsto f_1 \cdot_\hbar f_2 - f_2 \cdot_\hbar f_1$$

define, respectively, the structures of associative and Lie algebras on the classical function space $\mathcal{W}(\mathbf{R}^k \times \mathbf{R}^k)$.

To explore the dependence of the algebra defined by (17) on $\hbar$, one may proceed by expanding the exponential

(20) $$e^{(i\hbar/2)(a\cdot b'-a'\cdot b)}$$

as a series in powers of $\hbar$. From now on we shall assume that the classical function space is $\mathcal{S}(\mathbf{R}^k \times \mathbf{R}^k)$, the algebra (under both multiplication and convolution) of the *Schwartz functions* in q, p. The terms of the series can then be explicitly evaluated. For instance, the constant term is $(f_1 f_2)(q, p)$, while the next term becomes, after some calculation,

(21) $$\frac{i\hbar}{2}P(f_1, f_2)(q, p)$$

where $P(f_1, f_2) = \{f_1, f_2\}$ is the Poisson Bracket. Thus

(22) $$f_1 \cdot_\hbar f_2 = f_1 f_2 + \frac{i\hbar}{2}P(f_1, f_2) + O(\hbar^2).$$

We define the *Moyal Bracket* of f_1 and f_2 by

(23) $$[f_1, f_2]_\hbar = f_1 \cdot_\hbar f_2 - f_2 \cdot_\hbar f_1.$$

Then

(24) $$[f_1, f_2]_\hbar = i\hbar\{f_1, f_2\} + O(\hbar^2).$$

This is just the Dirac prescription but with the essential refinement that $[f_1, f_2]_\hbar$ *is not exactly equal to* $i\hbar\{f_1, f_2\}$ *but is only equal to it modulo* $\hbar^2$. We have thus constructed the quantum algebra associated to a general classical system, giving meaning *ex posto facto* to Dirac's profound and remarkable insight. We shall see presently that for the canonical variables the Dirac and Moyal brackets are the same, and the same is true for angular momentum observables and certain obvious generalizations of them.

A similar calculation gives us all the higher order terms as well. To make it transparent we introduce two sets of variables

$$q^{(1)}, p^{(1)}, q^{(2)}, p^{(2)}$$

and the associated Fourier transform variables

$$a^{(1)}, b^{(1)}, a^{(2)}, b^{(2)}.$$

We then have the differential operator

$$\Pi = \sum_{1 \le j \le k} \left(\frac{\partial}{\partial q_j^{(1)}} \otimes \frac{\partial}{\partial p_j^{(2)}} - \frac{\partial}{\partial p_j^{(1)}} \otimes \frac{\partial}{\partial q_j^{(2)}} \right)$$

and let $\widehat{\Pi}$ be the induced operator on the Fourier transform space via the transform. Then $\widehat{\Pi}$ is multiplication by the polynomial

$$\widehat{\Pi} = \sum_{1 \le j \le k} \left(-a_j^{(1)} b_j^{(2)} + b_j^{(1)} a_j^{(2)} \right). \tag{25}$$

Finally, let Res be the operator from $\mathcal{S}(\mathbf{R}^{2k} \times \mathbf{R}^{2k})$ to $\mathcal{S}(\mathbf{R}^{2k})$ which is the restriction to the diagonal:

$$\mathrm{Res}(f)(q, p) = f(q, p, q, p).$$

Then

$$P = \mathrm{Res} \circ \Pi(f_1 \otimes f_2).$$

We define the bidifferential operator P_N by

$$P_N : f_1, f_2 \longmapsto \left(\mathrm{Res} \circ \Pi^N\right)(f_1 \otimes f_2). \tag{26}$$

Then the general term of the expansion of (17) in powers of $\hbar$ is

$$\frac{(i\hbar/2)^N}{N!} P_N(f_1, f_2) \tag{27}$$

We thus have the formal expansion of (17):

$$[f_1, f_2]_\hbar = 2 \sum_{M=0}^{\infty} \frac{(i\hbar/2)^{2M+1}}{(2M+1)!} P_{2M+1}(f_1, f_2) \tag{28}$$

Symbolically we write this as

$$f_1 \cdot_\hbar f_2 = e^{(i\hbar/2)P}, \qquad [f_1, f_2]_\hbar = \frac{2}{\hbar} \sin \frac{\hbar}{2} P \tag{29}$$

which go back to Moyal's paper [26].

Let us summarize this discussion. Let $\mathcal{S}$ be the Schwartz space of the classical phase space $\mathbf{R}^{2k}$. Then $\mathcal{S}$ is a commutative algebra under multiplication. For any real value of the parameter $\hbar$, the definition

$$f_1 \cdot_\hbar f_2 = e^{(i\hbar/2)P}$$

interpreted by (17) converts $\mathcal{S}$ into an associative algebra denoted by $\mathcal{S}_\hbar$. Then $\mathcal{S}_\hbar$ is a family of noncommutative associative algebras which have $\mathcal{S}$ as their limit when $\hbar \to 0$. Moreover, under the commutator product, $\mathcal{S}_\hbar$ becomes a Lie algebra whose Lie bracket becomes the Poisson Bracket on $\mathcal{S}$ in the limit when $\hbar \to 0$:

$$[f_1, f_2]_\hbar := f_1 \cdot_\hbar f_2 - 2_1 \cdot_\hbar f_1 = i\hbar\{f_1, f_2\} + O(\hbar^3) \tag{30}$$

It is natural to ask if this product structure can be defined for larger algebras than $\mathcal{S}$. This can be done at least *formally* by treating $\hbar$ as a *formal parameter.* Let us write $t = i\hbar/2$, treating t as a formal parameter and define

$$f_1 \cdot_t f_2 = e^{tP} = f_1 f_2 + \sum_{r \geq 1} \frac{t^r}{r!} P_r(f_1, f_2), \tag{31}$$

where the series is treated formally. It then makes sense *even when the* f_i *are in* $C^\infty(\mathbf{R}^{2k})$ because the P_r are bidifferential operators. The associative law persists in the formal sense because the identities that have to be established for associativity are differential operator identities that are true for all differentiable functions once they are true for Schwartz functions. Thus (31)

defines the structure of an associative algebra in the formal sense on $C^\infty(\mathbf{R}^{2k})$. We shall make this more precise in the next section.

The quantum algebra as a limit of finite dimensional quantum algebras

This idea, as we have mentioned earlier, was due to Weyl and taken up by Schwinger later [6, 19]. Kinematically this was done in great generality in [20]. To understand the limiting process dynamically we consider, for simplicity, motion on a line where the potential is big at infinity (it is enough that $V(q)/\log|q| \to \infty$ as $|q| \to \infty$) so that the energy spectrum is discrete, as in the case of the harmonic oscillator. Indeed, taking N to be an odd natural number, one can identify $\mathbf{Z}_N$ with the finite grid $X(N) \subset \mathbf{R}$:

$$X(N) = \{r\varepsilon_N \mid r = 0, \pm 1, \pm 2, \ldots, \pm N^0\}$$

where

$$\varepsilon_N = \left(\frac{2\pi}{N}\right)^{1/2}, \quad N = 2N^o + 1.$$

The Hilbert space of the system is now $L^2(X(N))$; the position operator q_N is multiplication by the coordinate; the momentum operator p_N is the Fourier transform of q_N, the Fourier transform being the one defined for the finite group $\mathbf{Z}(N) = \mathbf{Z}\varepsilon_N/N\cdot\mathbf{Z}\varepsilon_N$ with which $X(N)$ can be identified. The Hamiltonian is

$$H_N = \frac{1}{2}p_N^2 + V(q_N).$$

In the limit when $N \to \infty$ this system goes over to the usual system on the real line with the Hamiltonian

$$H = \frac{1}{2}p^2 + V(q).$$

An invariant way to formulate this limit is to ask that

$$\mathrm{Tr}\big(f(H_N)\big) \longrightarrow \mathrm{Tr}\big(f(H)\big) \qquad (N \to \infty)$$

for a large class of functions f for which $f(H)$ is of trace class. See [6, 19, 20, 21] where this is done in arbitrary dimension and for functions of the form e^{-sx} of the real variable x, leading to the result that the dynamical operators in imaginary time converge in a very strong manner:

$$||\exp(-sH_N) - \exp(-sH)||_1 \to 0 \qquad (N \to \infty)$$

where $||\cdot||_1$ is the trace norm in the Banach space of trace class operators.

2.5 Quantum algebras over phase space

The Weyl–Moyal theory may be summarized by saying that quantum mechanics is essentially equivalent to imposing on the classical *commutative algebra* of functions on the phase space *a family of non commutative algebra structures* depending on $\hbar$ such that (22)–(24) are satisfied, thus giving the Dirac insight a very satisfactory interpretation. Depending upon the choice of the classical algebra, the quantum parameter $\hbar$ can be analytic or formal. The Hilbert space of quantum mechanics now arises as the space on which these non commutative algebras are represented by operators.

Mathematicians will recognize in the relations (22)–(31) the fact that the quantum algebra may be viewed as a *deformation* of the classical algebra with $\hbar$ as the deformation parameter. If I am not mistaken, this point of view was pioneered by (the late) Moshe Flato and his collaborators and colleagues, F. Bayen, C. Fronsdal, A. Lichnerowicz, D. Sternheimer, and H. Basart, in a series of seminal papers [27, 28, 29, 30]. Although deformations of associative and Lie algebras had been an established field, the application of this subject to the foundations of quantum theory was new and was first done by them. I shall now try to describe briefly some of the main results in this development which is still very active. This has fitted well with the recent fascination among theoretical physicists for non commutative geometry and so there is a great following for this type of work.

The calculations involving the Weyl–Moyal product and the Moyal bracket may be viewed as describing *a class of deformations of* $\mathcal{S}$ *or* $C^\infty(\boldsymbol{R}^{2k})$, *viewed as a Poisson algebra.* To explain this point of view let us begin with the notion of deformation of associative algebras first. We take a naive point of view.

Analytical deformations

Let $\mathcal{A}$ be an associative algebra over $\mathbf{C}$. Intuitively one speaks of a *deformation of* $\mathcal{A}$ when one is given a family of associative algebras $\mathcal{A}_t$ defined for small values of a parameter t whose underlying vector spaces coincide with

that of $\mathcal{A}$ such that item (i) if we write

$$f, g \mapsto f \cdot_t g$$

for the product in $\mathcal{A}_t$, then the product map is nice in some sense (smooth, analytic, etc)

(ii) For $t = 0$, $f \cdot_t g$ reduces to fg, the product in $\mathcal{A}$.

t is called the *deformation parameter*.

Formal deformations

Here one views t *as an indeterminate* and passes to $\mathcal{A}[[t]]$ which is the algebra of formal power series

$$f = a_0 + a_1 t + a_2 t^2 + \ldots \qquad (a_i \in \mathcal{A})$$

Clearly the product

$$f, g \mapsto fg$$

in $\mathcal{A}$ extends to a product

$$\sum f_p t^p, \sum g_q t^q \mapsto \sum k_r t^r$$

in $\mathcal{A}[[t]]$ where

$$k_r = \sum_{p,q \geq 0, p+q=r} f_p g_q$$

We denote by

$$\mathcal{A}[[t]]_0$$

the algebra thus obtained. A *formal deformation of* $\mathcal{A}$ is now defined as an associative algebra whose underlying space is $\mathcal{A}[[t]]$ and whose multiplication

$$f, g \mapsto f \cdot_t g \quad (f, g \in \mathcal{A}[[t]])$$

is such that when $t = 0$ it reduces to the algebra $\mathcal{A}[[t]]_0$. If we write

$$f \cdot_t g = fg + tC_1(f, g) + t^2 C_2(f, g) + \ldots \quad (f, g \in \mathcal{A})$$

where

$$C_j : \mathcal{A} \times \mathcal{A} \longrightarrow \mathcal{A} \qquad (C_0 = \text{id})$$

are bilinear maps, then the associativity requirement means that the C_j satisfy the identities

$$\sum_{r\geq 0, s\geq 0, r+s=n} C_r(a, C_s(b, c)) = \sum_{r\geq 0, s\geq 0, r+s=n} C_r(C_s(a, b), c)$$

Conversely, if the $C_j : \mathcal{A} \times \mathcal{A} \longrightarrow \mathcal{A}$ are bilinear maps satisfying these identities, the definition of product for two elements of $\mathcal{A}$ extends to an associative algebra structure on the vector space $\mathcal{A}[[t]]$ that coincides with the algebra structure $\mathcal{A}[[t]]_0$ when $t = 0$. The reason for introducing formal deformations is that they are algebraic objects and so may be studied with more ease than analytical deformations. Moreover, any analytical deformation in which the product is analytic in the parameter gives rise to a formal deformation. However *true quantization requires analytical deformations and not formal deformations, so that the formal deformations become more remote from the physical point of view.* In particular, if in any given context we establish the existence of a formal deformation, there still remains the problem of making analytical sense for these formal deformations for at least a large subalgebra of the original algebra, if the deformation is to be given a physical interpretation, for instance, to obtain a representation in a Hilbert space. We call such deformations *quantizations*.

Considerable work has been done in the directions suggested by the above remarks. Quantizations of symplectic and Poisson manifolds have been obtained. In the most important case of $\mathbf{R}^{2k}$ as the phase space it was shown by Moshe Flato and his collaborators that the Weyl-Moyal deformation is *essentially the only one*, thus exhibiting quantum mechanics as the *only possible deformation of classical mechanics*. In other words the Dirac quantum algebra is the only one possible. In my opinion this uniqueness theorem is the most fundamental foundational result in making transparent the transition from the classical to the quantum algebra.

2.6 Moshe Flato remembered

The idea that the Weyl-Moyal algebras should be treated in the broader framework of deformations of Poisson algebras was pioneered by Moshe Flato and his collaborators F. Bayen, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer. Soon others such as De Wilde, Le Conte, Fedesov, Deligne, and Kontsevich

took up this theme and went very far with it. It is unfortunate that Moshe passed away at the peak of his creative powers. I want to pen a few lines in which I try to give a glimpse of the mercurial personality of this remarkable man.

We first met in Liege in the mid 1970s in a beautiful conference on representation theory and fundamental physics that he organized. After that we always spent time together during his visits to UCLA, and once in Dijon when he invited me for a talk in his seminar. He told me once that had we met earlier we could have done much work together. When the AMS and the International Press published my Selected Papers, I sent him a copy with a letter recalling to him this remark. But unfortunately it reached him when he was already in a coma.

He had a very broad view of physics and expressed his preferences openly without worrying whether they were politically correct. He was that rare person who was very strict with the big people and very gentle with younger and less established scientists. To illustrate his kindness to young people the following incident would suffice. I had a student, Luca Guerrini, working on deformations of Lie algebras under me and after he finished his thesis I asked him to write up his results and submit it to the *Letters in Mathematical Physics*. The submission was rejected but I noticed in the rejection letter a tiny ray of hope: it was the comment that the paper does not explain why anyone would or should be interested in the questions studied in the submission. I told Luca that I will help him write an introduction and resubmit without any other change. This was done and the paper got accepted without requiring any modifications almost immediately. Some months later I was in Paris and met Moshe just for a few hours. The first thing he told me was "don't think I did not know that you wrote that intro." It turned out that this was the last time that I would see him for he died a few months later. A few months earlier he had agreed to be the mentor for Charlie Conley, another student of mine. Moshe suggested new lines of investigations and put Charlie in touch with people who were doing such things. This changed Charlie's entire scientific life in a dramatic manner and launched him on an extremely successful research career.

When Moshe passed away, there was going on a work shop on representation theory at UCLA, organized by Chris Fronsdal and Joe Wolf. I gave a eulogy for Moshe at that workshop, which was later published in one of the volumes of the Proceedings of the Conference that was organized in his memory (by Daniel Sternheimer and others). I reproduce it here.

Moshe Flato (1937–1998)[1]

Moshe Flato died on Friday, Nov 27, 1998, in Paris, of complications from a massive cerebral haemorrhage he suffered a couple of days earlier. His shocking and untimely death is a tremendous loss to the world of mathematical physics, and even more so, to his family and large circle of friends and admirers. He was buried in Tel-Aviv.

Moshe's work covers a large spectrum of high energy physics, straddling diverse aspects of quantum field theory and elementary particles. His interest was always in fundamental questions. He was constantly searching for a conceptual understanding of the strange world of elementary particles, and explored many new ideas and ways of thinking about them. He created a school in Dijon and Paris and shaped the intellectual developments of a huge number of young scientists, some of whom became his closest friends, such as Daniel Sternheimer and Jacques Simon.

I would like to make some very brief comments on his work. In quantum mechanics there is a famous uniqueness theorem due to von Neumann and Stone, which says that any system of operators satisfying the Heisenberg–Weyl commutation rules is isomorphic to the standard system which is called the Schrödinger model. The content of this theorem is that there is only one way to quantize a classical system with the specified commutation rules, namely through canonical quantization. In a famous paper published in the mid 1970's, Moshe, and his collaborators Bayen, Fronsdal, Lichnerowicz, and Sternheimer, proved a wonderful variant of this uniqueness in a very different context. Their idea was to think of quantization as a deformation of the Poisson algebra of classical observables. Canonical quantization furnishes such a deformation, the Moyal–Weyl deformation. They proved that this is essentially the only deformation of the classical Poisson algebra. In my opinion this uniqueness theorem lies at a much greater depth than the von Neumann–Stone theorem and may be thought of as the definitive clarification of the questions that were raised first by Hermann Weyl in 1927 regarding the foundations of quantum physics. The ideas coming out of this paper revolutionized quantization theory, made it possible to think of quantization in contexts where canonical quantization does not make sense (curved space-time, constrained mechanics, etc), and led inevitably to the theory of quantum

[1] Talk delivered on Dec 5, 1998, in the Lie group workshop held at UCLA on Dec 5–6, 1998.

groups and quantum homogeneous spaces. Recent work on q-quantum mechanics by Finkelstein and others would not have been possible without this seminal achievement. During the 1980's, Moshe, in collaboration with Fronsdal, made a deep analysis of the representation theory of the conformal group. Their theory of the so called singletons, which was very much ahead of its time, has recently come to the forefront in high energy physics.

Throughout his life he was deeply interested in formulating and understanding the nonlinear aspects of quantum field theory. He called this the "nonlinear program" and wrote many papers on this theme, with Simon and others. His recent work in collaboration with Simon and Taflin is a technical tour de force on the Cauchy problem and scattering theory for the coupled Maxwell and Dirac equations that arise in quantum electrodynamics.

He loved the subject of mathematical physics deeply and was very unhappy that (in those days) there was no forum where important and interesting discoveries in mathematical physics could be published quickly. The physicists had several such journals, but it was often very difficult to get papers published in them because of lack of sympathy of the physics community toward major discoveries that were more mathematical and did not lead to immediate physical interpretations. Moshe understood this situation and with characteristic energy and enthusiasm founded, in 1977, the *Letters in Mathematical Physics* precisely to correct this anamoly. The *Letters* has become one of the premier journals in mathematical physics. If you look through its pages you will find all the major developments in the last two decades represented there when they first came up. There is no question that he was the moving spirit behind the *Letters*. This is one achievement that will live long after him.

But bigger than all of his achievements in science is the fact that he was a decent man. Behind his rough exterior lay a heart of extraordinary generosity. He was lavish in giving his time to others. I learned a lot of physics by talking to him and reading his papers. He was especially generous to young people, understood their difficulties at the start of their scientific careers, and helped them in every possible way. He treated his friends royally, and influenced their lives in ways that cannot be enumerated. He was one of the few men I have known of whom it could be said that he was larger than life (a comment also made by one of the young scientists who was deeply influenced by Moshe).

As one gets older, one finds invariably that one cannot do all the things that one did when one was young, and certainly not at the depth that was reached in youth. The natural tendency to compensate for this is to reduce

one's ambitions and do what one can to advance science. Moshe scorned this way of living. He pushed himself and others around him without any slackening and lived on the creative edge all the time, right up to the moment of his death. According to the Hindu view of life this is one of the noblest ways to live and die.

When I learned of Moshe's death I was reminded of Isaac Stern's words on the occasion of David Oistrakh's death, when he referred to Oistrakh as a golden man. Moshe was a golden man.

References

[1a] P. A. M. Dirac, Proc. Roy. Soc. A 109 (1925), 642–653.

[1b] B. L. van der Waerden, *Sources of Quantum Mechanics*, Dover, 1967, 307–320; this source book will be referred to as SQM.

[2] W. Heisenberg, Zeit. Phys. 33 (1925), 879–893; see also SQM, 261–276.

[3] M. Born and P. Jordan, Zeit. Phys. 34 (1925), 858–888; SQM, 277–306.

[4] M. Born, W. Heisenberg, and P. Jordan, Zeit. Phys. 35 (1926), 557–615.

[5] SQM, 53–54, Letter to Pauli.

[6] H. Weyl, *Theory of Groups and Quantum Mechanics*, Dover, 1931.

Weyl's ideas were first published in H. Weyl, Zeit. Phys. 46 (1927), 1–46.

[7] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, 1955.

[8] V. S. Varadarajan, *Geometry of Quantum Theory*, Springer-Verlag, 1985.

[9] V. S. Varadarajan, Int. Jour. Theor. Phys. 32 (1993), 1815–1834.

[10] E. G. Beltrametti, and G. Cassinelli, *The Logic of Quantum Mechanics*, Encyclopedia of Mathematics and its Applications, Vol 15, Addison–Wesley, 1981.

[11] B. C. Van Fraassen, *Quantum Mechanics*, Oxford, 1991.

[12] G. Birkhoff and J. Von Neumann, Ann. Math. 37 (1936), 823–843.

[13] G. Cassinelli, E. De Vito, P. Lahti, and A. Levrero, Rev. Math. Phys. 9 (1997), 921–941.

[14] J. Schwinger, Proc. Nat. Acad. Sci. USA 45 (1959), 1542–1553; see also *Selected papers (1937–1976) of Julian Schwinger*, M. Flato, C. Fronsdal, K. A. Milton (eds), D. Reidel Publishing Company, Boston, 127–138.

See also the beautiful book of Schwinger, especially the Prologue:

J. Schwinger, *Quantum Mechanics: Symbolism of Atomic Measurements*, Ed. Berthold-Georg Englert, Springer, 2001.

[15] J. Schwinger, *Quantum Kinematics and Dynamics*, W. A. Benjamin, New York, 1970.

[16] J. Schwinger, *Hermann Weyl and quantum kinematics*, in *Exact Sciences and Their Philosophical Foundations*, Verlag Peter Lang, Frankfurt, 1988, 107–129.

[17] L. Accardi,Nuovo Cimento, 110 (1995), 685–721.

[18] G. W. Mackey, *Unitary group Representations in Physics, Probability, and Number theory*, Benjamin, 1978.

[19] J. Schwinger, Proc. Nat. Acad. Sci. USA, 46 (1960), 570–579.

[20] E. Husstad, *Endeligdimensjonale approksimasjoner til kvantesystemer*, Thesis, University of Trondheim, 1991/92.

[21] T. Digernes, V. S. Varadarajan, and S. R. S. Varadhan, Rev. Math. Phys. 6 (1994), 621–648.

[22] M. H. Stone, *Linear transformations in Hilbert Space. III. Operational methods and group theory*, Proc. Nat. Acad. Sci. USA, 16, (1930), 172–175.
[23] J. von Neumann, *Die Eindeutigkeit der Schrödingerschen Operatoren*, Math. Ann., 104 (1931), 570–578.
[24] V. S. Varadarajan, in *Analysis, Geometry, and probability,* Rajendra Bhatia (ed), Hindustan Book Agency, New Delhi, 1996, 362–396. For Mackey's paper see Duke Math. Jour. 16 (1949), 313–326, and also the book *Tata Lectures on Theta III* by D. Mumford with Madhav Nori and Peter Norman, Birkhäuser,Basal, 1991, pp. 1–14.
[25] V. S. Varadarajan, Lett. Math. Phys. 34 (1995), 319–326.
[26] J. E. Moyal, Proc. Camb. Phil. Soc. 45 (1949), 99–124.
[27] F. Bayen, M. Flato, M, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer. *Quantum mechanics as a deformation of classical mechanics*, Lett. Math. Phys. 1 (1975/77), no. 6, 521–530.
[28] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, *Deformation theory and quantization. I. Deformations of symplectic structures*, Ann. Physics 111 (1978), no. 1, 61–110.
[29] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, *Deformation theory and quantization. II. Physical applications*, Ann. Physics 111 (1978), no. 1, 111–151.
[30] H. Basart, M. Flato, A. Lichnerowicz, and D. Sternheimer, *Deformation theory applied to quantization and statistical physics*, Lett. Math. Phys. 8(1984), 483–494.

Chapter 3
Probability in the quantum world*

> When you watch you find that the electron goes either through one hole or the other, but if you are not looking you cannot say that it either goes one way or the other! Such is the logical tightrope on which Nature demands that we walk if we wish to describe her.
> *Richard Feynman*

Probability has been a part of quantum theory from its very beginnings. The original probabilistic interpretation of quantum mechanics was put forward by Max Born. However the idea that the understanding of Nature had to be statistical was unacceptable to a lot of people including Einstein, and there was substantial criticism of this aspect of quantum theory. Eventually, the ideas of von Neumann, Bell, Feynman, Mackey, Gleason, and many others on the probabilistic aspects of quantum theory clarified the situation and answered the criticisms. These contributions have made the role of probability in the quantum world both far-reaching and profound.

3.1. The statistical interpretation of quantum theory
3.2. The uncertainty principle of Heisenberg
3.3. Hidden variables
3.4. EPR
3.5. Transition probabilities in the atom
3.6. Feynman path integrals and the Feynman–Kac formula

3.1 The statistical interpretation of quantum theory

The most controversial part of quantum mechanics, at the time of its creation, was its statistical aspect. People were unwilling to accept that Nature allows us only to calculate the *probability distributions* of physical quantities. Of course one can measure any quantity (in principle) with no dispersion, but the theory insisted that we pay for it by having to forego any knowledge of certain other observables. There was also a feeling that the new ideas demanded the

* This essay and the previous one are based on lectures given at Howard University, Washington D.C., sponsored by my friend D. Sundararaman, in the 1990s.

abandoning of *causality*. The questions raised were deep and puzzling, and were not fully resolved for almost a decade or more. The roles of Einstein and Bohr in these discussions were decisive and are well documented.

The fundamental statistical interpretation of quantum theory is due to Max Born. If ψ is the wave function of the electron, he postulated that

$$\int_E |\psi(x, y, z)|^2 dxdydz$$

is the probability of finding the electron in the region $E \subset \mathbf{R}^3$. This is a special case of the following more general statement that was later obtained by von Neumann: if $A_i (1 \leq i \leq k)$ are commuting self adjoint operators representing simultaneously measurable observables, and P is their joint spectral measure, and if the system is in the state $[\theta]$ represented by the unit vector θ (we often say the state θ), the probability that a measurement on the A_i will yield a value of $(A_1, \ldots, A_k)$ belonging to a region $E \subset \mathbf{R}^k$ is

$$||P_E \theta||^2.$$

To see how the Born prescription is a special case of this all we have to do is to take the Hilbert space to be $L^2(\mathbf{R}^3)$ and the operator A_i to be the one that corresponds to the i^{th} coordinate of the location of the electron, namely, A_i is the multiplication by x_i. As another example, if A is a self adjoint operator with simple discrete spectrum $a_1, a_2, \ldots$ and corresponding normalized eigenvectors $\psi_1, \psi_2, \ldots$, then, for a system in state $[\theta]$ if a measurement is made to find the value of A, the probability that the value will be a_r is $|(\psi_r, \theta)|^2$ for $r = 1, 2, \ldots$. If $\mathcal{H}$ is the Hilbert space associated with a quantum system, and if ψ, θ are two unit vectors, the quantity

$$|(\psi, \theta)|^2$$

is the probability that if the state of the system is $[\theta]$ ($[\theta]$ is the state defined by θ) and an experiment is performed to verify that it is $[\psi]$, then the system will be found in the state $[\psi]$. In particular, the expectation value $\mu_\theta(A)$ and and the dispersion $\sigma_\theta^2(A)$ of A in the state $[\theta]$ are

$$\mu_\theta(A) = (A\theta, \theta), \qquad \sigma_\theta^2(A) = ((A - \mu_\theta(A)1)^2\theta, \theta)$$

In principle, *all* statistical assertions of quantum theory are derivable from these.

This prescription furnishes the details of the fundamental departure of quantum theory from classical theory, namely, the fact that in quantum theory *the state determines only the probability distributions of all physical quantities*. There is no abandonment of causality: the state itself moves causally as determined by the Schrödinger equation, but the theory denies the possibility of determining the exact values of all the observables in any state. It is very easy to give examples of states, especially in the finite models, in which certain observables have sharp values while others do not have this property. This was the source of a profound controversy about the consistency and completeness of the quantum mechanical view of nature which we shall analyze in more detail in a little while. For the moment we shall concentrate on the mathematical source of the statistical interpretation.

One way to understand this picture is through von Neumann's propositional logic of quantum mechanics, namely the logic $\mathcal{L}$ of closed subspaces of the Hilbert space $\mathcal{H}$. Here we think of $\mathcal{L}$ as a set with a partial order (implication) $\subset$, and an orthocomplementation (negation) $S \longmapsto S^{\perp}$. For example, suppose we are looking at one particle moving in $\mathbf{R}^3$ and the experiment is measuring its position at a certain time t, then the quantum proposition would be that the particle was found in a region $E \subset \mathbf{R}^3$. The Hilbert space is $L^2(\mathbf{R}^3)$, and the proposition is represented by the subspace of vectors vanishing outside E. If instead we speak of locating the momentum vector of the particle in a region $F \subset \mathbf{R}^3$, the corresponding subspace is the subspace of states *whose Fourier transforms vanish outside* F (Here Fourier transforms are in the L^2-sense).

In full generality, if S is a closed subspace of $\mathcal{H}$ representing an experimental proposition to be verified, then in the state ψ the probability that the experimental proposition is true is

$$||P_S\psi||^2$$

where P_S is the orthogonal projection on S. These probabilities completely determine the distribution of any observable in the state ψ; indeed, if A is any observable and P^A is its spectral resolution, then, in the state ψ the distribution of A is the probability measure

$$E \longmapsto ||P_E^A\psi||^2.$$

This shows that the assignment

$$p_\psi : S \longmapsto ||P_S\psi||^2$$

on the logic $\mathcal{L}$ of the system, the logic of all closed subspaces of $\mathcal{H}$, completely determines all the statistical properties of observables in the state ψ. The p_ψ are often called *vector states*. It is easily verified that p_ψ has the following properties:

(a) $0 \leq p_\psi(S) \leq 1$ for all $S \in \mathcal{L}$
(b) $p_\psi(0) = 1$, $p_\psi(\mathcal{H}) = 1$
(c) If S_i are mutually orthogonal closed subspaces of $\mathcal{H}$ and S is their orthogonal direct sum, then

$$p_\psi(S) = \sum_i p_\psi(S_i)$$

These properties generalize those of a probability measure on a Boolean σ-algebra and so one can think of p_ψ as a probability measure on the logic $\mathcal{L}$. In fact, we shall call any function $p(a \to p(a))$ on an arbitrary logic with these properties a *probability measure* on the logic. If p is any probability measure on $\mathcal{L}$, then for any observable A with spectral measure P^A we may define the probability distribution of A with respect to p as the probability measure

$$E \longmapsto p(S_E), \qquad S_E = \text{ the range of } P^A_E$$

So one can define a more general notion of a state of the system as a probability measure on $\mathcal{L}$; the earlier notion is obtained if this measure is of the form p_ψ. Note that the correspondence

$$[\psi] \longmapsto p_\psi$$

is bijective. Indeed, p_ψ depends only on $[\psi]$, and if $p_\psi = p_\theta$ then

$$0 = p_\psi(\psi^\perp) = p_\theta(\psi^\perp) \Longrightarrow \theta \in (\psi^\perp)^\perp \Longrightarrow [\theta] = [\psi]$$

Now the set of probability measures is a convex set closed under infinite convex combinations. These represent *mixed states* in the conventional theory and are appropriate for quantum statistical mechanics. If one considers only the convex combinations of the vector states, p_ψ it is an easy consequence that we obtain the measures of the form

$$p_T : S \longmapsto \text{Tr}\,(T P_S) = \text{Tr}\,(T^{1/2} P_S T^{1/2}),$$

where T is a self adjoint nonnegative operator of trace class and trace 1. In conventional quantum theory T is called a *density matrix* or the *density operator*; the measures p_T were first discovered by von Neumann, and (independently) by Landau. Such T have a discrete spectrum and if its nonzero eigenvalues are $c_i > 0$ with corresponding eigenvectors ψ_i (orthonormal), then

$$p_T = \sum_i c_i p_{\psi_i} \qquad (c_i > 0, \sum_i c_i = 1)$$

showing that if the system is in the state T, it is in one of the states $[\psi_i]$ with probability c_i, so that p_T is a *classical superposition* of the states $[\psi_i]$. If the eigenvalues are simple, this decomposition is unique, but when there are multiplicities, there is nonuniqueness since for a multiple eigenvalue the choice of the orthonormal basis in the corresponding eigenspace is not unique. However, T is uniquely determined by p_T.

Now von Neumann had to assume much more about states to be able to say that they must be of the form p_T. The question whether *any* measure on $\mathcal{L}$ is of the form p_T was raised by Mackey and was solved by Gleason in a famous paper [1, 2] where he showed that if $\dim(\mathcal{H}) \neq 2$, then every probability measure on $\mathcal{L}(\mathcal{H})$ is of the form p_T for a unique density operator T. The restriction on the dimension is essential. It can then be shown that the vector states p_ψ are the *extreme points* of the convex set of all states. Traditionally they are known as *pure states*; in the conventional model they are simply called the states, and represent the states with maximal information.

In dimension 2 there are bizarre states not coming from density operators T; however there is no point in considering them because as soon as one forms the composite of a two-dimensional system with another one, we are in the case allowed by Gleason's theorem, and so we are in the framework of states defined by density operators. Thus in all cases it is natural to work with the states p_T.

It is only after the Mackey–Gleason analysis that the conventional picture is complete, with the points of the projective space $P(\mathcal{H})$ of $\mathcal{H}$ representing the (pure) states of the system. The function

$$\pi : [\psi], [\theta] \longmapsto |(\psi, \theta)|^2$$

is well defined on $P(\mathcal{H}) \times P(\mathcal{H})$ and represents the probability structure on the set of states. A *symmetry* of the system is then a bijection of $P(\mathcal{H})$ that preserves π. The fundamental theorem of Wigner is that symmetries are

induced by unitary or antiunitary operators of $\mathcal{H}$ which moreover are unique up to a multiplicative phase factor (complex constant of absolute value 1). If the symmetry group is connected, the symmetries are unitary. For many connected groups the ambiguities involved in the choice of the phase factors in the unitaries implementing the symmetries can be resolved, so that we can say that as a rule, symmetry is governed by unitary representations of the symmetry group or some central extension of it by a torus. Thus taking the group to be the group of time translations, time evolution is governed by a unitary representation of $\mathbf{R}$, written

$$t \longmapsto U_t = e^{-\frac{it}{\hbar}H}$$

where H, a self adjoint operator whose existence and uniqueness is guaranteed by Stone's famous theorem on unitary one-parameter groups, is identified with the energy or the Hamiltonian of the system. How do the states evolve in time? Under time evolution given by the unitary group above the pure states move by

$$t \longmapsto \psi_t = e^{-\frac{it}{\hbar}H}\psi_0$$

where ψ_0 is the state at time $t = 0$; infinitesimally, differentiating with respect to t, this becomes

$$\frac{d\psi_t}{dt} = -\frac{i}{\hbar}H\psi_t$$

which is the *Schrödinger equation*. The mixed states p_T evolve by

$$T \longmapsto e^{-\frac{it}{\hbar}H}Te^{\frac{it}{\hbar}H}$$

The conventional physical interpretation of the statistics of the observables under time evolution is now immediate. One remark: the dynamical groups defined by H and $H + a$ where a is a real constant, induce the same actions on the projective space of $\mathcal{H}$. Thus the Hamiltonian is ambiguous up to an additive constant.

3.2 The uncertainty principle of Heisenberg

From the very beginning quantum mechanics faced serious problems of physical interpretation. The transition from phase space dynamics to the dynamics of non commutative algebras of observables and their probability distributions

raised difficult questions about physical interpretation. It was Heisenberg who took the crucial step of exploring the physical consequences of the commutation rules that dominated the early formulations of the theory. By a brilliant analysis based on thought experiments he was led to his famous uncertainty principle which has become the corner stone of all physical interpretations of quantum theory.

The *uncertainty principle* of Heisenberg illustrates the basic point that in quantum mechanics it is impossible that all observables have sharply defined values in any given state. For any observable A let μ_ψ and σ_ψ^2 denote its mean and dispersion in the state p_ψ. If σ_ψ^2 is small, then most of the probability distribution is concentrated in a small neighborhood of μ_ψ; in such a case we shall say that the observable is highly localized in the state ψ. Let us now consider the quantum system of a particle moving in $\mathbf{R}$ whose position and momentum operators are defined as multiplication by x and $-i\hbar d/dx$ on $\mathcal{H} = L^2(\mathbf{R})$. If ψ is a state to which these operators can be applied in succession (for instance if ψ is a Schwartz function), then, writing Δ_x^2 and Δ_p^2 for the respective dispersions of position and momentum, one can show that

$$\Delta_x \Delta_p \geq \frac{\hbar}{2}$$

which shows that any attempt to define the position sharply implies a very high dispersion in the value of the momentum and vice versa. This is the famous uncertainty principle of Heisenberg. This of course is mathematically easy to prove, and one can even give a complete description of the states where the product of dispersions reaches its minimum value, the so-called *coherent states*. However Heisenberg's real accomplishment was not the mathematical part but the physical part, wherein, using remarkable thought experiments that involve trying to locate an electron using a gamma ray microscope, he showed (see [3, 4]) that *the laws of classical electrodynamics and the wave–particle duality relations already place such severe restrictions on the errors of observations on the position and momentum that one must have, up to second order in* $\hbar$,

$$\Delta_x \Delta_p \simeq \hbar.$$

The wave-particle duality of light which figures prominently in Heisenberg's discussion was emphasized by Bohr as a corner stone of quantum theory. Bohr would eventually call it the *principle of complementarity* of quantum systems. Its basis is the fact, established by many experiments,

that all particles have wave properties. The photoelectric effect reveals really particle properties of light, while the phenomenon of electron diffraction can be explained only under the assumption that electrons have wave properties. The particle and wave aspects constitute the so-called *wave–particle duality of all matter*, a basic fact in atomic physics, that was first put forward by de Broglie. However the wave and particle aspects of matter are *complementary*; no single experiment can reveal both properties. This was how Bohr formulated his principle of complementarity. From our mathematical point of view, the non commutative nature of the logic reflects this complementarity.

Let me briefly discuss the mathematical part of the Heisenberg uncertainty principle. We have two self adjoint operators A, B satisfying the commutation rule

$$[A, B] = ia1 \qquad (a \in \mathbf{R}, a \neq 0)$$

We wish to show that if ψ is in the domains of both AB and BA,

$$\sigma_\psi(A)\sigma_\psi(B) \geq \frac{|a|}{2}$$

We replace A, B respectively by

$$A' = A - \mu_\psi(A)1, \qquad B' = B - \mu_\psi(B)1,$$

a change that does not affect the commutation rule. Then, from

$$(A'B'\psi - B'A'\psi, \psi) = ia$$

we get

$$(A'\psi, B'\psi) - (B'\psi, A'\psi) = 2i\Im(A'\psi, B'\psi) = -ia$$

Then, by Schwartz's inequality, writing $|a| = \varepsilon a$,

(*)

$$\frac{|a|}{2} = -\varepsilon\Im(A'\psi, B'\psi) \leq |(A'\psi, B'\psi)| \leq ||A'\psi||||B'\psi|| = \sigma_\psi(A)\sigma_\psi(B)$$

The same argument proves a more general result due to H. P. Robertson: if

$$[A, B] = iC$$

then

$$\sigma_\psi(A)\sigma_\psi(B) \geq \frac{1}{2}|(C\psi, \psi)|$$

As an illustration of this more general inequality Robertson considers the angular momenta

$$M_x, M_y, M_z$$

around the three axes which satisfy the commutation rules

$$[M_x, M_y] = i\hbar M_z, \text{ etc}$$

showing that *the product of the standard deviations of two components of the angular momentum vector in any state is always greater than or equal to* $\hbar/2$ *times the absolute value of the mean of the third component in that state.*

In the finite models of Weyl and Schwinger which we have talked about earlier, there are complementary observables x, y with the following property: if one of them has a sharp value, all the values of the other are equally probable.

3.3 Hidden variables

Classical statistical mechanics is also a statistical theory in which the states are probability distributions on phase space which evolve causally according to the equations of motion. From the beginning a major controversy had raged about the true nature of the statistical aspects of quantum theory. Many people felt that the reason why quantum theory has to be statistical is that one cannot know the exact state of the atom, so that the situation is very similar to statistical mechanics where averagings over the phase space are used because we cannot know the precise mechanical state of a system whose number of degrees of freedom is really huge ($\simeq 10^{23}$). The argument was then made that the situation is analogous in quantum mechanics and that the statistical properties of quantum systems are the result of averaging over some unknown or *hidden variables,* as they came to be called. This argument, thus pursued, then led to the conclusion that quantum mechanics was an *incomplete description of physical reality.* Whether this is actually true or not is the so-called *hidden variables question.*

The first to deal with this question in depth was von Neumann. He reached the remarkable conclusion that the results of quantum theory cannot be duplicated by any hidden variables mechanism. Because of the fundamental nature of the controversy and its importance, von Neumann gave both a mathematical and physical argument; they were not complementary but rather they

reinforced each other. For additional details see [2, 6] and my survey [5c] of Mackey's work.

The physical side of von Neumann's argument is as follows. If a hidden variables explanation is to be allowed, then we should be able to resolve any ensemble into finer and finer ones where more and more observables have sharply defined values. But this method of successive resolution is doomed from the start because of the fact that in atomic systems, any measurement changes the system, and hence changes the values of quantities which have been previously measured. Thus, in the resolution of atomic ensembles *we do not get ahead* by such a procedure and there seems no way of reaching those ensembles where every observable has a sharply defined value, ensembles that are called *dispersion free*. The analysis of Stern–Gerlach experiments is a vivid illustration of this argument.

Nevertheless, one may still cling to the hope that there may be other ways of reaching the dispersion-free ensembles. To examine this further it becomes necessary to enquire more into the structure of the dispersion-free ensembles and see whether their existence is in conflict with quantum mechanical experience. This leads us to the mathematical side of the argument.

The mathematical argument that I shall present now, which is based on the Mackey-Gleason analysis and is a significant sharpening of von Neumann's, is that *in the conventional model there are no states that are dispersion free except when we are in the trivial case where* $\dim(\mathcal{H}) = 1$. For obtaining this result in as general a context as possible von Neumann replaced the ensembles by the functionals characteristic of the ensemble, which associate to each bounded observable its expectation value in the state defined by the ensemble. So von Neumann viewed an ensemble as the functional

$$\mathcal{E} : \mathcal{B}_0 \longrightarrow \mathbf{R}$$

where $\mathcal{B}_0$ is the space of all bounded self adjoint linear operators on the Hilbert space $\mathcal{H}$, such that, for any $A \in \mathcal{B}_0$, $\mathcal{E}(A)$ is the expectation value of the observable represented by A in the state (=ensemble) identified with $\mathcal{E}$. It is natural to assume that this functional had the following properties:

(a) $\mathcal{E}(1) = 1$
(b) $\mathcal{E}(A^2) \geq 0$ for all $A \in \mathcal{B}_0$
(c) $\mathcal{E}(aA + bB) = a\mathcal{E}(A) + b\mathcal{E}(B)$ if A, B are *commuting elements* of $\mathcal{B}_0$ and $a, b \in \mathbf{R}$.

But von Neumann assumed much more; for instance (c) was required even when A and B do not commute, a much more tenuous hypothesis than (c) obviously. He also assumed, for technical reasons, continuity properties for $\mathcal{E}$. We shall not assume these extra properties. We shall call as *states* the functionals satisfying *only* the conditions (a)-(c). This definition of a state is the widest possible compatible with von Neumann's idea of identifying the ensemble with the state.

If $\mathcal{E}$ is a state and A is an observable, then the dispersion of A in that state is defined as the number

$$\mathcal{E}(A^2) - (\mathcal{E}(A))^2$$

which is always ≥ 0. The state is called *dispersion-free* if this is 0 *for all* A. In such a state, any observable A has a sharply defined value, namely $\mathcal{E}(A)$.

Let $\mathcal{L}$ be the logic of a physical system. For any $a \in \mathcal{L}$ let Q_a be the observable which is 1 if the experimental proposition a is verified to be true, and 0 if it is verified to be not true. The Q_a are the so-called *questions* of Mackey [5]. Then in any state there is a probability $p(a)$ that a is true,

$$\Pr(Q_a) = p(a).$$

Clearly

$$0 \leq p(a) \leq 1, \quad p(0) = 0, \quad p(1) = 1$$

and since Q_a takes only the values 0 and 1 we have

$$\text{Exp}(Q_a) = p(a), \qquad \text{Dispersion}(Q_a) = p(a)(1 - p(a))$$

So a state determines a function

$$p : \mathcal{L} \longrightarrow [0, 1], \qquad a \longmapsto p(a)$$

This is just the state functional but it is now restricted only to the questions. If $a_i \in \mathcal{L} (i = 1, 2)$ and $a_1 < a_2^{\perp}$ then one can form the sum $a = a_1 + a_2$ in $\mathcal{L}$, the two propositions are simultaneously verifiable, i.e., define commuting projections, and one should require that

$$p(a) = p(a_1) + p(a_2).$$

In other words, the function p is a *finitely additive probability measure* on $\mathcal{L}$ and every expectation functional gives rise to a finitely additive probability measure p on $\mathcal{L}$. If the state is dispersion-free, then $p(a)(1 - p(a)) = 0$

so that $p(a) = 0$ or 1 for any a. Let us call a finitely additive probability measure on $\mathcal{L}$ which takes only the values 0 and 1 a *two-valued measure. To prove that there are no dispersion-free states it is thus enough to prove that there are no two-valued measures on* $\mathcal{L}$.

Theorem. (*a*) *If* $\mathcal{L}$ *is a Boolean algebra, there are two-valued measures on* $\mathcal{L}$ *and every finitely additive probability measure on* $\mathcal{L}$ *is a natural mixture of these.* (*b*) *If* $\mathcal{L}$ *admits no two-valued measures, then* $\mathcal{L}$ *has no imbedding into a Boolean algebra.* (*c*) *If* $\mathcal{L}$ *contains a sublogic isomorphic to* $\mathcal{L}(\boldsymbol{D}^3)$ *where* $\boldsymbol{D} = \boldsymbol{R}, \boldsymbol{C}$, *then* $\mathcal{L}$ *admits no two-valued measures. In particular* $\mathcal{L}(\mathcal{H})$ *has no two-valued measures and so has no dispersion-free states if* $3 \leq \dim(\mathcal{H}) \leq \infty$.

Proof. (a) is just the classical theorem of Stone. In fact, $\mathcal{L}$ (is isomorphic to) is a Boolean algebra of subsets of some space X; if $x \in X$, the Dirac measure δ_x, which is one if $A \in \mathcal{L}$ contains x and 0 otherwise, is a two-valued measure. Clearly every measure is a mixture of these. (This argument reverses the historical development. Indeed, Stone proved the isomorphism of an abstract Boolean algebra with a Boolean algebra of sets in some space by first proving that there are two-valued measures, and identifying these as the points of his space!)

(b) follows trivially from (a).

To prove (c) it is enough to prove that there are no two-valued measures on $\mathcal{L}(\mathbf{D}^3)$. But if p is a two-valued measure, it is now *countably additive* and so Gleason's theorem applies and shows that p comes from a density operator T. If ψ is a unit vector, then $(T\psi, \psi) = p(\mathbf{D}\cdot\psi)$ is either 0 or 1, and so by continuity, is $\equiv 0$ or $\equiv 1$, showing that $T = 0$ or $T = 1$. This is impossible. If $\dim(\mathcal{H}) \geq 3$, then we can find a subspace S of dimension 3, and the map

$$S' \longmapsto S' \oplus S^{\perp} \qquad (S' \subset S)$$

is an imbedding of $\mathcal{L}(S)$ as a sublogic of $\mathcal{L}(\mathcal{H})$. This finishes the proof. □

The restriction on dimension is essential. It is possible to show that for the conventional model in dimension 2 the statistics can be obtained through the hidden variable mechanism (see [6]). This is somewhat remarkable and paradoxical because it was the Stern–Gerlach experiments involving spin that played a prominent role in understanding and motivating the conventional model!

The structures of measures on a Boolean algebra and a non commutative logic are thus essentially different. For instance, in the conventional model for quantum mechanics, not only are there no two-valued measures but *every probability measure on the logic takes all the values between* 0 *and* 1. The complementarity structure is too rich to allow gaps in the values. Moreover, in almost all models, the logic is a lattice, and for any measure p one has

$$p(a_1) = 1,\ p(a_2) = 1 \Longrightarrow p(a_1 \cap a_2) = 1$$

which is very surprising since the assumptions on p involve only commuting projections. The point is that the very fact that the measure be defined for *all* elements of the logic forces the coherence properties such as the ones above.

There are variations of the above argument where the emphasis is on formulating the hidden variables question as that of *imbedding the quantum propositional logic inside a Boolean algebra.* See

- N. Zierler, and M. Schlessinger, *Boolean imbeddings of orthomodular sets and quantum logic*, Duke Math. Jour., 32(1965), 251–262.
- S. Kochen, and E. P. Specker, *The problem of hidden variables in quantum mechanics*, J. Math. Mech., 17(1967), 331–348.

Imbeddings of the logic inside a Boolean algebra will give rise to two-valued measures on the logic which can be forbidden under suitable conditions. Thus, ultimately, it is the absence of two-valued measures on logics that is the reason for the impossibility of a hidden variables explanation of quantum mechanics.

This brief discussion does not do full justice to the remarkable contributions of many people to this circle of ideas, to the structure of quantum logics, to probability on logics, and the philosophical implications of these ideas. Some of these can be found in the references below. See also the articles in the Wheeler–Zurek [7b] and Lahti–Mittelstaedt [10c] volumes.

- J. M. Jauch, *Foundations of Quantum Mechanics*, Addison–Wesley, 1968.
- C. Piron, *Foundations of Quantum Physics*, Benjamin, 1976.
- M. Jammer, *The Philosophy of Quantum Mechanics*, Wiley, 1974.
- C. A. Hooker (Ed.), *The Logico-Algebraic Approach to Quantum Mechanics*, Vol. 1., Reidel, 1975.
- E. G. Beltrametti, and B. C. van Frassen, *Current issues in Quantum Logic*, Plenum, 1981.

Bell's inequalities

In the 1960's J. S. Bell, dissatisfied with the von Neumann argumentation, challenged the assumptions underlying it [7a]. It is difficult to accept Bell's objections in the *general framework* treated by von Neumann, Mackey and Gleason that we are considering. Nevertheless Bell's work was important because, by considering certain restricted situations, he was able *to bring the hidden variables question into the domain of experimental testing.*

Bell considers a system that consists of a pair of spin 1/2 observables which have interacted in the past and are now sufficiently apart to be distinguished as particles 1 and 2. Such systems had been considered earlier in connection with the so-called EPR paradox (see below). The spin orientations of the two particles can then be measured by Stern–Gerlach magnetic filters and their correlations computed and verified. One can also work with photons and their linear polarization states and measure these by polarization analyzers—the mathematics of computing the correlations between the two particles is essentially the same in both cases. Bell now wanted to see if these quantum correlations could be explained on the basis of a hidden variables model which was *local* in the sense that measurements can be made on each particle without apparent disturbance of the other. He was then led to what are now known as *Bell's inequalities* which had to be satisfied by the correlations if the statistics of the experiment that Bell proposed were to be explained by a local hidden variable mechanism. This discussion allowed him to propose experimental tests whether his inequalities were indeed violated in practice. Many experiments have since been conducted and they have revealed that Bell's inequalities are indeed violated, and that the observed correlations are exactly as predicted by quantum mechanics [8, 9]. I shall now discuss briefly Bell's ideas.

We begin with the two-dimensional Hilbert space $\mathcal{K} = \mathbf{C}^2$ and the usual model for quantum mechanics associated with it; the calculations below are standard. Let Σ be the set of obsevables which only take the values ± 1; these are parametrized by hermitian matrices $X(\mathbf{a})$ corresponding to unit vectors $\mathbf{a}$ given by

$$X(\mathbf{a}) = \begin{pmatrix} a_1 & a_2 + ia_3 \\ a_2 - ia_3 & -a_1 \end{pmatrix} \qquad \mathbf{a} = (a_1, a_2, a_3), \|\mathbf{a}\| = 1.$$

We choose normalized eigenvectors $s_\pm(\mathbf{a})$ of $X(\mathbf{a})$ for the eigenvalues ± 1. If $P_\pm(\mathbf{a})$ are the orthogonal projections on the one-dimensional spans of

$s_\pm(\mathbf{a})$, then

$$X(\mathbf{a}) = P_+(\mathbf{a}) - P_-(\mathbf{a}), \qquad I = P_+(\mathbf{a}) + P_-(\mathbf{a})$$

from which we get

$$P_\pm(\mathbf{a}) = \frac{I \pm X(\mathbf{a})}{2}.$$

We write $\mathbf{u}^*$ for the vector whose components are complex conjugates of those of the vector $\mathbf{u} \in \mathcal{K}$. If $\mathbf{u}$, $\mathbf{v}$ are two unit vectors and $P_\mathbf{u}$, $P_\mathbf{v}$ are the orthogonal projections on the one-dimensional spans of $\mathbf{u}$, $\mathbf{v}$, then

$$\mathrm{Tr}(P_\mathbf{u} P_\mathbf{v}) = |(\mathbf{u}, \mathbf{v})|^2.$$

Hence, as $\mathrm{Tr}(X(\mathbf{a})X(\mathbf{b})) = 2\,\mathbf{a}\cdot\mathbf{b}$, we get,

$$|(s_\pm(\mathbf{a}), s_\pm(\mathbf{b}))|^2 = \frac{1}{4}\mathrm{Tr}\Big((I \pm X(\mathbf{a}))(I \pm X(\mathbf{b})\Big) = \frac{1}{2}\Big(1 + (\pm)(\pm)\mathbf{a}\cdot\mathbf{b}\Big).$$

We shall now consider the state of the form

$$\tau = \frac{1}{\sqrt{2}}(\mathbf{t}_+^* \otimes \mathbf{t}_+ + \mathbf{t}_-^* \otimes \mathbf{t}_-)$$

where $\{\mathbf{t}_+, \mathbf{t}_-\}$ is an ON basis of $\mathcal{K}$. It is a simple calculation to verify the following crucial property of τ:

τ is independent of the choice of the ON basis $\{\mathbf{t}_+, \mathbf{t}_-\}$.

This means that for *any* ON basis $\{\mathbf{u}, \mathbf{v}\}$ for $\mathcal{K}$ we can write

$$\tau = \frac{1}{\sqrt{2}}\Big(\mathbf{u}^* \otimes \mathbf{u} + \mathbf{v}^* \otimes \mathbf{v}\Big)$$

Write

$$X_1(\mathbf{a}) = X(\mathbf{a}) \otimes I, \qquad X_2(\mathbf{a}) = I \otimes X(\mathbf{a}).$$

Then $X_1(\mathbf{a})$ and $X_2(\mathbf{b})$ commute for all $\mathbf{a}$, $\mathbf{b}$ and we can calculate their individual and joint distributions in the state τ. We use the above expression for τ when $\mathbf{u} = s_\pm(\mathbf{a})^*$, $\mathbf{v} = s_\mp(\mathbf{a})^*$ and $\mathbf{v} = s_\pm(\mathbf{b})$, $\mathbf{v} = s_\mp(\mathbf{b})$. First we have

$$P_\tau(X_1(\mathbf{a}) = \pm 1) = P_\tau(X_2(\mathbf{b}) = \pm 1) = \frac{1}{2}.$$

If we set

$$P_{\pm\pm}(\mathbf{a}, \mathbf{b}) = P_\tau(X_1(\mathbf{a}) = \pm 1, X_2(\mathbf{b}) = \pm 1)$$

then

$$P_{\pm\pm}(\mathbf{a}, \mathbf{b}) = |(\tau, s_\pm(\mathbf{a}) \otimes s_\pm(\mathbf{b}))|^2.$$

Now, if we write

$$\mathbf{a}^\sim = (a_1, a_2, -a_3)$$

then

$$X(\mathbf{a}^\sim) = X(\mathbf{a})^*$$

so that

$$X(\mathbf{a}^\sim) s_\pm(\mathbf{a})^* = \pm s_\pm(\mathbf{a})^*.$$

We may thus replace $s_\pm(\mathbf{a})^*$ by $s_\pm(\mathbf{a}^\sim)$. Hence

$$P_{++}(\mathbf{a}, \mathbf{b}) = \frac{1}{2}|(s_+(\mathbf{a}^\sim), s_+(\mathbf{b}))|^2 = \frac{1 + \mathbf{a}^\sim \cdot \mathbf{b}}{4}$$

$$P_{--}(\mathbf{a}, \mathbf{b}) = \frac{1}{2}|(s_-(\mathbf{a}^\sim), s_-(\mathbf{b}))|^2 = \frac{1 + \mathbf{a}^\sim \cdot \mathbf{b}}{4}$$

$$P_{+-}(\mathbf{a}, \mathbf{b}) = \frac{1}{2}|(s_+(\mathbf{a}^\sim), s_-(\mathbf{b}))|^2 = \frac{1 - \mathbf{a}^\sim \cdot \mathbf{b}}{4}$$

$$P_{-+}(\mathbf{a}, \mathbf{b}) = \frac{1}{2}|(s_-(\mathbf{a}^\sim), s_+(\mathbf{b}))|^2 = \frac{1 - \mathbf{a}^\sim \cdot \mathbf{b}}{4}$$

Let us write

$$\Gamma_{QM}(\mathbf{a}, \mathbf{b}) := E_\tau\Big(X_1(\mathbf{a}^\sim) X_2(\mathbf{b})\Big).$$

We then find

$$\Gamma_{QM}(\mathbf{a}, \mathbf{b}) = \mathbf{a} \cdot \mathbf{b}.$$

At the same time we have the remarkable fact that the value of $X_1(\mathbf{a}^\sim)$ *determines* the value of $X_2(\mathbf{a})$ and vice versa. This is the essence of the EPR property of the state τ which we shall discuss in the next paragraph. This property is simply the relation

$$P_\tau\Big(X_1(\mathbf{a}^\sim) = X_2(\mathbf{a})\Big) = 1$$

which is immediate from the formulae for $P_{\pm\pm}$ (we must remember that $\mathbf{a} \cdot \mathbf{a} = 1$).

Bell now asks the question *whether the correlations Γ_{QM} can be obtained from a local hidden variables model.* This means there should be a probability

space Ω with a probability measure m on it, and corresponding to each unit vector $\mathbf{a} \in \mathbf{R}^3$ two random variables $\xi(\mathbf{a}), \eta(\mathbf{a})$ such that $(\xi(\mathbf{a}), \eta(\mathbf{b}))$ has the same joint distribution as $(X_1(\mathbf{a}), X_2(\mathbf{b}))$. Since $X_1(\mathbf{a}^{\sim}) = X_2(\mathbf{a})$ with probability 1, it follows that we may assume that

$$\eta(\mathbf{a}) = \xi(\mathbf{a}^{\sim}).$$

If we define

$$\Gamma_{HV}(\mathbf{a}, \mathbf{b}) := E_m(\xi(\mathbf{a}^{\sim})\eta(\mathbf{b})) = E_m(\xi(\mathbf{a}^{\sim})\xi(\mathbf{b}^{\sim}))$$

then Bell's question is the following: is it possible to choose the probability measure m and the correspondence $X \longleftrightarrow \xi$ such that

$$\Gamma_{QM}(\mathbf{a}, \mathbf{b}) = \Gamma_{HV}(\mathbf{a}, \mathbf{b})$$

for all unit vectors $\mathbf{a}, \mathbf{b} \in \mathbf{R}^3$?

Bell answered this in the negative. He discovered that the correlations of *any set* of ± 1-valued random variables with mean 0 on a classical probability space necessarily satisfy some inequalities that are now called the *Bell's inequalities*. Clearly $|\xi_1 \xi_2| \leq 1$ so that $|E_m(\xi_1, \xi_2)| \leq 1$, but much more is true. Indeed, if $\xi_i (i = 1, 2, 3)$ are *any* three such variables on (Ω, m), then

$$1 + \xi_1\xi_2 + \xi_2\xi_3 + \xi_3\xi_1 \geq 0 \qquad \text{on } \Omega$$

because at any point of Ω two of the three variables must have the same sign. Replacing ξ by $\pm\xi$ and taking expectation values and writing

$$\gamma_{ij} = \gamma_{ji} = \Gamma_{HV}(\xi(\mathbf{a}_i^{\sim}), \xi(\mathbf{a}_j^{\sim}))$$

we have,

$$1 + \varepsilon_1\varepsilon_2\gamma_{12} + \varepsilon_2\varepsilon_3\gamma_{23} + \varepsilon_3\varepsilon_1\gamma_{31} \geq 0$$

where $\varepsilon_j = \pm 1$. Choosing the ε_j appropriately we get

$$|\gamma_{12} - \gamma_{13}| \leq 1 - \gamma_{23} \qquad (B_3)$$

To violate (B_3) we take the $\mathbf{a}_i$ in the plane $a_3 = 0$ (so that $\mathbf{a}_i = \mathbf{a}_i^{\sim}$) with angles $\pi/3$ between adjacent ones so that, writing $X_i = X(\mathbf{a}_i)$,

$$(X_1, X_2) = 1/2, \quad (X_2, X_3) = 1/2, \quad (X_3, X_1) = -1/2$$

Then, writing

$$\Gamma_{ij} = \Gamma_{QM}(X_i, X_j)$$

we have

$$|\Gamma_{12} - \Gamma_{13}| = 1 > |1 - \Gamma_{23}| = 1/2$$

For a detailed and beautiful discussion of the experiments to test Bell's inequalities see the account by Aspect and Grangier [8, 9]. For the original papers of Bell and others not only on this question but about the entire theory of measurement in quantum theory, see the monumental book of Wheeler-Zurek [7b].

In conclusion one can now say with a great amount of conviction that the hidden variables question has been decisively answered in the negative. *There is no way, compatible with the current experimental and theoretical structure of quantum theory, to obtain the results of quantum theory by any kind of reasonable classical hidden variable scheme.*

3.4 EPR

Can quantum mechanical description of physical reality be considered complete? This was the title of the famous paper [10a] of Einstein, Podolsky, and Rosen. In the initial years after the discovery of quantum mechanics and its probabilistic interpretation by Heisenberg, Dirac, Born, Jordan, and Schrödinger, Einstein was very much opposed to it and believed that it was not *internally consistent*. In the Solvay congress of 1930 he came up with a thought experiment in which in which a box containing radiation was weighed before and after the release of a single photon at a pre-assigned time. He then claimed that this would allow the determination of the energy of the photon at a precise time, thus violating the uncertainty relation between energy and time. Bohr refuted this by showing that according to Einstein's general relativity, the gravitational field involved in the weighing of the box will change the clock mechanism (that was present to release the photon at a precise time) *just enough* to maintain the uncertainty relation between energy and time. After this Einstein accepted the consistency of quantum theory. But in 1935 he published the paper referred above, in collaboration with Podolsky and Rosen, in which they purported to show the *incompleteness* of the quantum mechanical description of reality.

In this paper they constructed an example on the basis of which they argued that the quantum mechanical description of physical reality was necessarily

incomplete. Their example was a quantum system consisting of two particles which had interacted in the past but are now spatially far apart, and the state of the composite system had the following remarkable property: if the value of the position or momentum of the first particle is known, the value of the position or momentum of the second particle can be predicted with certainty. In this manner they argued that *without disturbing the second system* one can measure both the position and momentum of the second particle, a circumstance forbidden by quantum mechanics. From this they argued that that quantum theory was an *incomplete* description of what happens in the atomic world.

The impact of this paper on Bohr was remarkable. Here is what Rosenfeld, who was collaborating with Bohr at that time had to say. ... *The onslaught came down upon us like a bolt from the blue. Its effect on Bohr was remarkable. We were then in the midst of groping attempts at exploring the implications of the fluctuations of charge and current distributions, which presented us with riddles of a kind we had not met in electrodynamics. A new worry could not come at a less propitious time. Yet, as soon as Bohr had heard my report of Einstein's argument, everything else was abandoned: we had to clear up such a misunderstanding at once. We should reply by taking up the same example and showing the right way to speak about it...* [11a]

Bohr's reply was published in the same journal and had the same title as the EPR paper. By discussing explicit experimental arrangements Bohr argued that the phrase *without disturbing the second system* had an essential ambiguity because of the complementarity phenomena in quantum mechanics [11a, 11b].

The essence of the EPR example is the following property of the joint system: there is a special state of the composite system such that to each observable A in the second system one can associate an observable B in the first system with the property that if a measurement yields a sharp value of B, then one can predict the value of A with certainty. This of course is not possible in every state of the composite system. For position and momentum measurements of particles in one dimension the EPR state was the distribution state (of infinite norm necessarily since we are dealing with continuous spectra)

$$\sigma = \int e^{i(x_1 - x_2)p} dp \qquad (\hbar = 1)$$

The integrals have to be interpreted as tempered distributions. If X_i, P_i are the position and momentum observables of the i^{th} particle, then if a measurement of P_1 is made and yields the value p in the state σ, the state collapses to the state with wave function

$$e^{i(x_1-x_2)p}$$

This is an eigenstate of $-i\partial/\partial x_2$ for the eigenvalue $-p$. Thus in the state σ we have

$$P_1 \otimes 1 = p \Longrightarrow 1 \otimes P_2 = -p$$

On the other hand, a simple Fourier analysis shows that σ has another representation:

$$\sigma = \int \delta_x \otimes \delta_x dx$$

from which we also have, by a similar argument,

$$X_1 \otimes 1 = x \Longrightarrow 1 \otimes X_2 = x$$

Thus the knowledge of either the position or momentum of the first particle allows us to predict with certainty the position and momentum values of the second particle. It is not without interest to explore the ideas of this argument more carefully. I do not wish to add to Bohr's discussion and refer the reader to his article [10b, 11b]. The point roughly is that even though we do not seem to disturb the second system, we do disturb the first and hence the composite system. The spatial apartness of the two systems is a misnomer; any measurement of the first system immediately determines the value of a corresponding quantity in the second. The experiments are decisive on this point. See [9].

It is however of some interest to examine the structure of states like the ones that Einstein et all discovered, to characterize them, and preferably, to construct them with finite norm. We shall begin by extracting from the above discussion a very general definition of states with the characteristic property discovered by Einstein et al, at least for states of finite norm. Let us consider two systems S_i with Hilbert spaces $\mathcal{H}_i (i = 1, 2)$ and the composite system $S = S_1 \times S_2$ with Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Let $\sigma \in \mathcal{H}$ be a unit vector.

Definition. The state $[\sigma]$ is said to be an EPR state of S relative to a set Σ_2 of bounded observables of S_2 if to each $A_2 \in \Sigma_2$ there is a bounded observable A_1 of S_1 and a borel function g from $\mathbf{R}$ to $\mathbf{R}$ such that

$$P\left(1 \otimes A_2 = g(a) \mid A_1 \otimes 1 = a\right) = 1$$

for all $a \in \mathbf{R}$; A_2 is then said to be linked to A_1. If Σ_2 is the set of all bounded observables of S_2, we shall say that σ is EPR with respect to S_2.

In order to determine when such states exist and to determine all of them, it is convenient to use a special description of $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. We use the well-known fact that the elements of $\mathcal{H}$ are in natural one-one correspondence with *conjugate linear* operators

$$L : \mathcal{H}_2 \longrightarrow \mathcal{H}_1$$

such that

$$\mathrm{Tr}(L^\dagger L) < \infty$$

where $L^\dagger(\mathcal{H}_1 \longrightarrow \mathcal{H}_2)$ is the adjoint of L. The condition on L is equivalent to requiring that

$$\sum_n ||Le_n||^2 < \infty$$

for some ON basis (e_n) of $\mathcal{H}_2$, and then

$$\mathrm{Tr}(L^\dagger L) = \sum_n ||Le_n||^2$$

for *any* ON basis (e_n) of $\mathcal{H}_2$. The corresponding element of $\mathcal{H}$ is

$$\sum_n Le_n \otimes e_n$$

and is independent of the choice of the (e_n). We thus identify the states of S with the operators L.

The key fact is that if L is any state of S, it is EPR with respect to Σ_2 if and only if $\Sigma_2 \subset \mathcal{B}_L$, the commuting algebra of $L^\dagger L$, so that $\mathcal{B}_L$ is the largest set with respect to which L is EPR. If now L is to be EPR with respect to S_2 it is necessary and sufficient that $L^\dagger L$ is a scalar; this is the same as saying that

$$\dim(\mathcal{H}_2) < \infty \text{ and } L = \dim(\mathcal{H}_2)^{-1} U$$

with

$$U : \mathcal{H}_2 \hookrightarrow \mathcal{H}_1 \text{ an anti-unitary injection .}$$

The set of EPR states for S_2 is then in canonical bijection with the set of anti-unitary injections of $\mathcal{H}_2$ inside $\mathcal{H}_1$.

It is now easy to see that if the EPR state is defined by the anti-unitary injection $U(\mathcal{H}_2 \hookrightarrow \mathcal{H}_1)$, then, for any observable B of S_2, UBU^{-1} is the observable of S_1 linked to B by the EPR property. Intuitively, the existence of such an imbedding of the observable algebra of S_2 inside that of S_1 is reasonable because the propositions of S_2 must be found within S_1. By Wigner's theorem such an imbedding should be implemented by a symmetry $\mathcal{H}_2 \hookrightarrow \mathcal{H}_1$; the technical points that go beyond this heuristic reasoning are that $\mathcal{H}_2$ must be finite dimensional and the symmetry should be *anti-unitary*[12a].

Let us now finally assume that both Hilbert spaces are of the same finite dimension d. Then we conclude that the states that are EPR for S_2 are also EPR for all observables of S_1 also, and are precisely all states of the form

$$\sigma = d^{-1/2} \sum_{1 \le j \le d} Ue_j \otimes e_j$$

where U is an *anti-unitary isomorphism* of $\mathcal{H}_2$ with $\mathcal{H}_1$, and (e_j) is *any* ON basis of $\mathcal{H}_2$. The map

$$B_2 \longleftrightarrow B_1 = UB_2U^{-1}$$

then has the property that the value of either of the B_i determines the value of the other with certainty. The state τ in the discussion of Bell's inequalities is an example of such an EPR state where the anti-unitary isomorphism $\mathcal{K} \longrightarrow \mathcal{K}$ is just complex conjugation. Examples of states such as these were first constructed by Bohm [12b]. Thus finally we see that all EPR states have the same form as the states that Bohm worked with [12a].

We have mentioned already that one can construct the analogues of the original Einstein–Podolsky–Rosen states in certain finite quantum systems. We take $\mathcal{H}_1 = \mathcal{H}_2 = L^2(G)$ where G is a finite abelian group of order $|G|$. The scalar product is given by

$$(f, g) = \frac{1}{|G|} \sum_{x \in G} f(x)g(x)^{\text{conj}} \qquad (f, g \in L^2(G))$$

For simplicity we consider the antiunitary isomorphism $U : f \longmapsto f^{\text{conj}}$ of $\mathcal{H}_2$ with $\mathcal{H}_1$. If (e_n) is any ON basis of $\mathcal{H}_1$ we have the representation of the corresponding state σ_U as

$$\sigma_U = d^{-1/2} \sum e_n^{\text{conj}} \otimes e_n$$

Now we have two ON bases of $\mathcal{H}_2$, namely $\{\sqrt{|G|}\delta_x\}_{x\in G}$ and $\{\xi\}_{\xi\in\widehat{G}}$ where δ_x is the delta function at x and $\widehat{G}$ is the group of characters of G. So

$$(*) \qquad \sigma_U = \sqrt{|G|} \sum_{x\in G} \delta_x \otimes \delta_x = \frac{1}{\sqrt{|G|}} \sum_{\xi\in\widehat{G}} \xi^{-1} \otimes \xi$$

Let X_2 be a "position" observable in S_2 with distinct values $a_x (x \in G)$ and corresponding eigenstates δ_x. For Y_2 we take the "momentum" observable in S_2 with distinct values b_ξ and eigenstates ξ. Let $X_1 = UX_2U^{-1}$, $Y_1 = UY_2U^{-1}$. For the pair $(\mathcal{H}_1, \mathcal{H}_2)$ the EPR relations are summarized by the symmetrical formulae

$$P(X_1 \otimes 1 = a_x, 1 \otimes X_2 = a_x) = 1, \quad P(Y_1 \otimes 1 = b_\xi, 1 \otimes Y_2 = b_\xi) = 1.$$

If we now compare these formulae with those of Einstein et al we see that we have here the finite analogue of their state. The original EPR state can then be realized as certain renormalized limits of these finite model states, in their continuum limit [12].

3.5 Transition probabilities in the atom

The quantum theory developed in the 1920s had been able to solve, at least in principle, all problems concerning atoms and molecules. However the relationship of this theory to the origins of quantum mechanics was not explained at all. In particular, the Einstein–Bohr frequency relations and the probabilities of the atomic transitions, which were worked out by Einstein based on thermodynamic considerations *before* the development of quantum theory, were not part of the new theory. This was not surprising since these features flowed from the interaction of the atom with the electromagnetic field, and this had not been studied.

Dirac was the first to treat the problem of deriving these aspects of quantum theory from the basic principles of quantum mechanics. In a famous and beautiful paper [13, 14] written in 1927 (see also the treatment of the Dirac theory by von Neumann [4] and Weyl [15]), Dirac treated this problem by considering the quantum system consisting of the atom and the *quantized ra-*

diation field, and wrote down a suitable Hamiltonian for the combined system on the basis of which he calculated the transition probabilities between the energy levels of the atom; the theory led him to the Eisnstein–Bohr frequency relations as well as to the Einstein formulae for the transition probabilities.

Dirac's theory was quite remarkable and introduced many features that were absolutely new. First of all he had to face the fact that the classical radiation field was a system with *infinitely* many degrees of freedom and so its quantization went well beyond the principles of quantization of systems with finite number of degrees of freedom that had been treated hitherto. Moreover the quantized radiation field had to be given a particle interpretation right from the outset because the Einstein–Bohr relations involved the emission and absorption of *photons*, which meant that the particle aspect of radiation had to be at the very heart of the treatment. In Dirac's theory it came out *automatically* that the classical radiation field, which was a wave field, acquired particle properties upon quantization, and that the particles obeyed the Bose–Einstein statistics, as they should because they were photons. This was the *wave-particle duality of light*, finally emerging out of Dirac's treatment. Dirac's theory was the first quantum theoretic treatment of a field, although it was non relativistic.

I shall not go into the details of the Dirac theory except to give the formulae for the transition probabilities in unit time. In these formulae $I(\nu)$ is the intensity of the radiation field at the frequency ν, and k, k' respectively refer to the initial and final states of the atom and ν_k, $\nu_{k'}$ their respective frequencies. The factors $|\mathbf{Q}_{k'k}|^2$ were originally conjectured by Heisenberg. For absorption we have

$$\nu_{k'} > \nu_k, \ \nu = \nu_{k'} - \nu_k \qquad P_a \approx \frac{4\pi^2}{3\hbar^2}|\mathbf{Q}_{k'k}|^2 I(\nu)$$

and for emission

$$\nu_k > \nu_{k'}, \ \nu = \nu_k - \nu_{k'}$$

$$P_e \approx \frac{4\pi^2}{3\hbar^2}|\mathbf{Q}_{k'k}|^2\left[I(\nu) + \frac{\hbar\nu^3}{\pi^2 c^3}\right]$$

It is to be noted that the probability of emission is the sum of two terms one of which depends on the intensity of the radiation field exactly as in the case of the absorption case, while the other term is *independent of the intensity of the radiation field.* This phenomenon had already been discovered by Einstein who called it *spontaneous emission.*

Planck's radiation formula

Suppose now that there are many atoms in the radiation field and the entire system is in equilibrium. Then, on the average, the number of transitions $\nu_k \to \nu_{k'}$ will be the same as the number of transitions $\nu_{k'} \to \nu_k$. Let N_k be the number of atoms of energy $E_k = \hbar\nu_k$. Then [15] the equation of balance leads to

$$\frac{N_k}{N_{k'}} = 1 + \frac{i(\nu)}{I(\nu)}$$

where

$$i(\nu) = \frac{\nu^3\hbar}{\pi^2 c^3}$$

On the other hand, by Boltzmann's theory,

$$N_k = \text{ constant } e^{-E_k/KT} = \text{ constant } e^{-\hbar\nu_k/KT}$$

from which we get

$$e^{-\hbar(\nu_k - \nu_{k'}/KT} = 1 + \frac{i(\nu}{I(\nu)} \qquad (\nu = \nu_{k'} - \nu_k > 0)$$

This leads to the formula

$$I(\nu) = \frac{\hbar\nu^3}{\pi^2 c^3 (e^{\hbar\nu/KT} - 1)}$$

which is Planck's radiation law!

3.6 Feynman path integrals and the Feynman–Kac formula

In 1942, Richard Feynman, a graduate student in Princeton University, wrote a dissertation in physics that contained a novel way of formulating quantum mechanics. Because of the war the publication of the results of this thesis was delayed and it was not until 1948 that they first appeared in print [17, 14]. His method was to start with a classical system and associate to it some sort of an integral on the space of classical trajectories using which he gave prescriptions for calculating the probabilities of events in the associated quantum system. This integral was completely determined by the classical action functional. This method, nowadays called the method of *path integrals*, has since then

spread into all aspects of quantum physics, and for many physicists it is the preferred mode of thinking and calculation. I shall now give a brief discussion of this method, the difficulties in formulating it as a rigorous mathematical instrument, and some situations where this can be done.

The heuristic thinking behind Feynman's method is as follows. In classical mechanics, if we consider two points x, y in configuration space, then in general there is only one path the particle can follow in going from x to y in time t. This path is in fact the one that makes the action a minimum, or at least an extremum. But in quantum mechanics, because of Heisenberg's uncertainty principle, the position and momentum cannot both be exactly determined at any instant of time, so that the particle is smeared out and one can no longer say what its path is with certainty; the double slit experiments demonstrate this brilliantly. Feynman's idea was that *one should consider all possible paths that go from x to y in time t and average over them to get the quantum probability of going from x to y in time t, but that this averaging should not be with respect to a conventional integral because of the wave–particle duality and the uncertainty principle.* He had the remarkable insight that *each path* from x to y in time t has an associated *complex amplitude, and that it is the average of these amplitudes over the space of paths that gives the complex amplitude whose modulus squared represents the probability that we want.* Inspired by a famous paper [16, 14] of Dirac on the Lagrangian in quantum mechanics, Feynman then made his final and most beautiful discovery that the amplitude associated to a path u is proportional to

$$e^{(i/\hbar)S[u]}$$

where S is the classical action. Using this more or less as a hypothesis he showed how to get all the basic results of quantum mechanics. His method, based on the Lagrangian rather than the Hamiltonian, proved to be very supple and would eventually be carried over to the context of quantum fields. For Feynman's beautiful discussions see [17, 18].

The amplitude for getting from x to y in time t is called the *propagator* and is denoted by $K_t(x, y)$. One way to think of this is to regard the unitary operator U_t that takes one from time $t = 0$ to time t as an integral operator with kernel $K_t(x, y)$. Feynman's heuristic argument leads to the following formula, called the *Feynman path integral formula*, for the propagator:

$$K_t(x, y) = \int e^{(i/\hbar)S[u]} \mathcal{D}u$$

where $\mathcal{D}u$ is some kind of a "Lebesgue" measure on the space of paths u such that $u(0) = x$, $u(t) = y$.

Unfortunately it is impossible to find a candidate for the "Lebesgue measure" $\mathcal{D}u$ that takes the place of summation in this formula. The attempts going back to Feynman to define the integral representation of the propagator by a limiting process similar to the ones used in the theory of Riemann integration are not convincing from the mathematical point of view because one cannot define a theory of integration that way. *The integration theory has to be defined first, and then one can use such approximation schemes to compute a well defined integral, but not the other way around.* Nevertheless this method has proved very popular among physicists. From the mathematician's point of view the method itself has to be thought of as a formal instrument with great suggestive power, which can often lead to predictions that can be verified by other methods.

This brief discussion does not do justice to the beauty and novelty of Feynman's ideas. In [18] he gives a brilliant discussion of the familiar double-slit experiment with electrons passing through two holes and falling on a screen, and how one is led to the uncertainty principle of Heisenberg. One can perform the experiment with very weak streams of electrons and ask what are the probabilities P, P_1, P_2 where P is the probability that the electron arrives at a position x on the screen when both holes are open, and P_i is the probability when only hole i is open ($i = 1, 2$); these are functions of x. If these events have the meaning they have classically, we would have

$$P = P_1 + P_2$$

but experimentally we find that

$$P \neq P_1 + P_2.$$

The interpretation is that when both holes are open, we cannot say that the electron goes through one hole or the other in order to arrive at the screen! To avoid making this interpretation the answer is to make an experiment to make sure that the electron does indeed pass through one or the other hole, by inserting a light source at each hole which would scatter the electron when it passes through that hole. The results are unambiguous: every electron that arrives at the screen has passed through one or the other hole! In Feynman's heuristic picture, the events have probability amplitudes (which are complex

vectors) ϕ, ϕ_1, ϕ_2 and one always has $\phi = \phi_1 + \phi_2$ but for the probabilities we may have $P = |\phi|^2 \neq P_1 + P_2 = |\phi_1|^2 + |\phi_2|^2$. I give here an extended quotation from [18]:

> ... We thus see that any physical agency designed to determine through which hole the electron passes, must produce, lest we have a paradox, enough disturbance to alter the distribution...
>
> It was first noticed by Heisenberg, and stated in his uncertainty principle, that the consistency of the then new mechanics required a limitation to the subtlety to which experiments could be performed. In our case it says that an attempt to design apparatus to determine through which hole the electron has passed and delicate enough so as not to deflect the electron sufficiently to destroy the interference pattern, must fail. It is clear that the consistency of quantum mechanics requires that it must be a general statement involving all the agencies of the physical world which might be used to determine through which hole the electron passes. The world cannot be half quantum mechanical and half classical. No exception to the uncertainty principle has been discovered....
>
> ... To summarize then: the probability of an event (in an ideal experiment where there are no uncertain external disturbances)is the absolute square of a complex quantity called the probability amplitude. When the event can occur in several alternative ways the probability amplitude is the sum of the probability amplitude for each alternative considered separately.
>
> If an experiment capable of determining which alternative is actually taken is performed the interference is lost and the probability becomes the sum of the probability for each alternative.

The precise *mathematical formulation* of the Fenyman path integral, at least in the case of particles moving in euclidean space under some otential field (and later, for some quantum fields) did not come till the 1950s and was a consequence of some deep ideas of Mark Kac [22b].

Feynman–Kac formula

In the 1950's the mathematician Mark Kac noticed that if we change in the Schrödinger equation the time variable t to it where $i = (-1)^{1/2}$ then it becomes a diffusion equation which governs the transition probabilities of a *Markov process*. To any stochastic process one can always associate a probability measure on the space of paths of the process and so one has a genuine path integral for computing the propagator of this process. This path integral can then be analytically continued in the time variable and the Feynman propagator can be obtained by this analytic continuation. The formula that Kac obtained by pursuing this idea is known as the *Feynman–Kac formula* and is one of the landmarks of rigorous path integral theory [19, 20]. I shall now discuss briefly this formula. In particular the derivation of this formula will

"explain" at least partially Feynman's profound insight that the functional inside the integration for the propagator is proportional to the exponential of the action.

To motivate the formula let us go to the finite model where a particle moves in a finite configuration space X, and the Hilbert space is $\mathcal{H}$, the space of complex functions on X.[1] Let us assume that the Hamiltonian of the quantum system is of the form

$$H = H_0 + V, \qquad H_0 = -\Delta_X$$

where $H_0 = -\Delta_X$ is the "free" part and V is a "potential", V being a real function on X. The notation is clearly set up to highlight the analogy with the quantum mechanics of a particle in $\mathbf{R}^k$ moving in the field of a potential function V; there the Hamiltonian is $-\Delta_k + V$ where Δ_k is the Laplacian in $\mathbf{R}^k$. The dynamical group is $U_t = e^{itH}$ and we wish to get a path integral representation for the kernel $K_t(x, y) = (U_t\delta_y, \delta_x)(x, y \in X)$. Because of the finite dimensionality of $\mathcal{H}$ the *complex dynamical group* $U_z = e^{zH}(z \in \mathbf{C})$ makes sense and so one can define $T_t = e^{-tH}$. In the infinite dimensional case this argument does not apply but if we assume that $H \geq 0$ in the spectral sense, then U_z is defined for $\Re(z) \leq 0$ without any problem. However, unlike the finite case T_t is defined only for $t > 0$; *semigroups rather than groups in the infinite dimensional case*.

The basic assumption now is that the free part H_0 has the following properties:

1. $\Delta_X 1 = 0$
2. If f is a real function on X and x_0 is a point at which f is minimum, then $(\Delta_X f)(x_0) \geq 0$

The second condition is known as the *minimum value principle* for Δ. This has a clear analogy with the Laplacian. Let

$$a(x, y) = (\Delta_X\delta_y, \delta_x), \qquad A = (a(x, y))_{x,y\in X}$$

so that A is the matrix of Δ_X in the basis (δ_x). It can be shown that these conditions are equivalent to saying that the semigroup $e^{-tH_0} = e^{t\Delta_X}$ is *Markovian*, i.e., the function

$$p_t(x, y) = (e^{t\Delta_X}\delta_y, \delta_x) \qquad (t > 0)$$

[1] I owe this discussion to my friend Varadhan.

is the transition probability of a Markov particle being at y at time t if at time 0 it was at x. This is the same as requiring that

$$p_t(x, y) \geq 0 \qquad \sum_y p_t(x, y) = 1$$

The above conditions are also equivalent to the following.

1. $\sum_y a(x, y) = 0$
2. (2) $a(x, y) \geq 0$ $(x \neq y)$ $(\Longrightarrow a(x, x) \leq 0)$

One can then build a family

$$P_x(x \in X)$$

of *genuine* probability measures on the space $\mathcal{F}$ of all step functions ω on the positive time axis $[0, \infty)$ which are right continuous and take values in X such that the probability

$$P_x\left(\omega \mid \omega(t_1) = x_1, \omega(t_2) = x_2, \ldots, \omega(t_N) = x_N\right)$$

is equal to

$$p_{t_1}(x, x_1) p_{t_2 - t_1}(x_1, x_2) \ldots p_{t_N - t_{N-1}}(x_{N-1}, x_N)$$

We shall not go into the construction of these measures, although even in this simple situation this is somewhat nontrivial.

Given the measures P_x we can define the conditional probability measures $P_{t,x,y}$ by

$$P_{t,x,y}(E) = P\left(\omega \in E \mid \omega(t) = y\right)$$

We then have the Feynman–Kac formula in this situation:

$$K_{it}(x, y) = \int_{\mathcal{F}} e^{-\int_0^t V(u(s))ds} \, dP_{t,x,y} \cdot p_t(x, y)$$

The original propagator itself is obtained by analytic continuation of the imaginary time propagator constructed above.

To prove this we use the so-called *Trotter product formula* which is trivial in the finite dimensional case but requires more care in the infinite dimensional case:

$$e^{t(\Delta_X - V)} = \lim_{N \to \infty} \left(e^{t\Delta_X/N} e^{-tV/N}\right)^N$$

A simple calculation shows that the matrix of

$$\left(e^{t\Delta_X/N}e^{-tV/N}\right)^N$$

is

$$\sum p_{t/N}(x, z_1)p_{t/N}(z_1, z_2)\dots p_{t/N}(z_{N-1}, y)e^{-(t/N)[V(z_1)+\dots+V(z_N)]}$$

where the sum is over all possible $z_1, z_2, \dots, z_{N-1}$. Now the expression

$$p_{t/N}(x, z_1)p_{t/N}(z_1, z_2)\dots p_{t/N}(z_{N-1}, z_N)$$

can be written as

$$P_{t,x,y}\left(\omega(t/N) = z_1, \dots, \omega((N-1)t/N) = z_{N-1}\right) \cdot p_t(x, y)$$

and so the matrix of

$$\left(e^{t\Delta_X/N}e^{-tV/N}\right)^N$$

becomes

$$\int e^{-(t/N)\sum_{1\le r\le N} V(\omega(rt/N))} dP_{t,x,y}(\omega) \cdot p_t(x, y)$$

The exponent is a Riemann sum which, when $N \to \infty$, goes to

$$\int_0^t V(\omega(s))ds$$

and so the whole expression has the limit

$$\int_{\mathcal{F}} e^{-\int_0^t V(\omega(s))ds}\, dP_{t,x,y}(\omega) \cdot p_t(x, y)$$

which is the formula of Feynman–Kac in this very special case.

Let us now go over from this "toy" situation to the continuous case when the configuration space is $\mathbf{R}$. The Hamiltonian is

$$H = -\frac{1}{2}\frac{d^2}{dx^2} + V(x)$$

where we assume that the potential is ≥ 0. It is known that if V is continuous this is an essentially self adjoint operator on the space of smooth functions with compact support and is ≥ 0. Hence we can speak of the semigroup $e^{-tH}(t \geq 0)$. By analogy with what we did in the finite case we must asso-

ciate a probability measure with the semigroup $e^{(t/2)d^2/dx^2}$ $(t \geq 0)$. We can use Fourier transform theory to give sense to this operator. At the Fourier transform level d^2/dx^2 is multiplication by $-\xi^2$ and so $e^{(t/2)d^2/dx^2}$ is multiplication by $e^{-(t/2)\xi^2}$; in the original space it is then *convolution* by the inverse Fourier transform of $e^{-(t/2)\xi^2}$ which is the gaussian density

$$\frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$$

So $e^{(t/2)d^2/dx^2}$ $(t \geq 0)$ is an integral operator with the gaussian kernel

$$g_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t}$$

This is the transition probability density of the well known *Wiener process* which models brownian motion. The corresponding probability measures W_x defined on the space $C[0, \infty)$ of all continuous functions on $[0, \infty)$ and model *brownian motion* starting from x at time 0. It is Markovian and in fact has the property of having independent increments, i.e., the increments

$$x(t_1) - x(s_1), \ldots, x(t_n) - s(s_n)$$

are independent if the intervals $[s_1, t_1], \ldots [s_n, t_n]$ are disjoint. It is not surprising that the free part of the Hamiltonian corresponds to brownian motion because in the approximating finite case it is a random walk and typically the random walks have the browninan motion as their limits. The final formula of Feynman–Kac is

$$\text{(FK)} \qquad K_{it}(x, y) = \int_{C[0,t]} e^{-\int_0^t V(u(s))ds} \, dW_{t,x,y} \cdot g_t(x, y)$$

Here $W_{t,x,y}$ are the conditional brownian motion measures that start from x at time 0 and reach y at time t. The meaning of the propagator is as follows:

$$e^{[(t/2)d^2/dx^2 - V(x)]}$$

is an integral operator on $L^2(\mathbf{R})$ with kernel $K_{it}(x, y)$. Formally we get this by replacing

$$e^{i\int_0^t (1/2)\dot{x}(s)^2} \mathcal{D}x$$

by the (conditional) Wiener measures in Feynman's formula. The real time propagator is obtained by analytic continuation [22c].

In the F-K formula the potential is *separated from the free part*, and the integration is determined by a measure that is independent of the potential and involves only the free part. None of these is obvious from the Feynman heuristics and so the Feynman–Kac formula is indeed a great advance beyond the original formula of Feynman.

How can this formula be used to get physically interesting results? One of the best situations where this formula can be applied is when we wish to obtain *spectral information* on the energy H. For this it makes no difference whether we consider e^{itH} or e^{-tH}. For a thorough treatment of many such applications see [19, 20]. One of the striking results is that if an atom has a ground state it is unique.

There are many generalizations of this formula. For systems with internal symmetry where the wave function has several components and the potentials are hermitian matrix-valued, see [21a, 21b]. There are however even more fundamental directions in which the theory can be extended, for instance, working in a Riemannian manifold where there is a natural Brownian motion corresponding to the Laplace-Beltrami operator. For homogeneous spaces such processes go back to the work of Yosida, but the probabilistic theory is governed by the great insights of Ito, and, after him, of Stroock and Varadhan. The theory has applications that reach deep into analysis and has allowed people (Bismut, Gestler, etc) to obtain proofs of the Atiyah-Singer index theorem. The literature is huge. It appears that the probabilistic method, by being able to follow what happens when one moves along a path of the Brownian motion, is able to really reach deep into the manifold. It is perhaps not an exaggeration to view it as a theory of the quantized manifold.

Kac himself considers the FK-formula as one of his most striking achievements. Here is his assessment taken from his autobiography [22a].

> In its various guises the F–K formula is ubiquitous throughout much of quantum physics on the one hand and probability theory on the other. It is probably safe to say that I am better and more widely known for being the K in the F–K formula than for anything else I have done during my scientific career.
>
> But first a little history. At the end of the war Hans Bethe returned to Cornell from Los Alamos and brought with him a group of brilliant young experimental and theoretical physicists. Among them was Richard Feynman who, to no one's surprise, went on to become a leading physicist of our day. It must have been in the spring of 1947 that Feynman gave a talk at the Cornell Physics Colloquium based on some material from his 1942 Ph. D dissertation, which had not yet been published.
>
> A fundamental concept of quantum mechanics is a quantity called the propagator, and the standard way of finding it (in the nonrelativistic case) is by solving the Schrödinger equation. Feynman found another way based on what became known

as the Feynman path integral or "the sum over histories" which Dyson alludes to in his book *Disturbing the universe.*

> During his lecture Feynman sketched the derivation of his formula and I was struck by the similarity of his steps to those I had encountered in my work. In a few days I had my version of the formula, although it took some time to complete a rigorous proof. My formula connected solutions of certain differential equations closely related to the Schrödinger equation with Wiener integrals.
>
> It is only fair to say that I had Wiener's shoulders to stand on. Feynman, as in everything else he has done, stood on his own, a trick of intellectual contortion that he alone is capable of.
>
> I find Feynman's formula to be very beautiful. It connects the quantum mechanical propagator, which is a twentieth century concept, with the classical mechanics of Newton and Lagrange, in a uniquely compelling way.. . .

References

[1] A. M. Gleason, *Measures on the closed subspaces of a Hilbert space*, J. Math. Mech. 6 (1957), 885–893.

[2] V. S. Varadarajan, *Geometry of Quantum Theory*, Second Edition. Springer-Verlag, New York, 1985.

[3] W. Heisenberg, *The Physical Principles of Quantum Theory*, Dover, 1930.

[4] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, 1955.Translated by R. T. Bayer from Von Neumann's 1932 Springer monograph *Mathematische Grundlagen der Quantenmechanik*.

[5a] G. W. Mackey, *Quantum mechanics and Hilbert space*, Amer. Math. Monthly 64(1957) no. 8, part II, 45–57. See also

[5b] G. W. Mackey, *The Mathematical Foundations of Quantum Mechanics*: W., A. Benjamin, 1963. For a detailed retrospective review of Mackey's work, see

[5c] V. S. Varadarajan, *George Mackey and his work on representation theory and foundations of physics*, in *Group representations, ergodic theory, and mathematical physics: a tribute to George W. Mackey*, 417–446, Contemp. Math., 449, (Eds.) R. S. Dornan, C. C. Moore, and R. J. Zimmer. Amer. Math. Soc., 2008.

[6] E. G. Beltrametti, and G. Cassinelli, *The Logic of Quantum Mechanics*, Encyclopedia of Mathematics and its Applications, Vol 15, Addison–Wesley, 1981.

[7a] J. S. Bell, *On the problem of hidden variables in quantum mechanics*, Rev. Modern Phys., 38(1966), 447–452. For a collection of papers on quantum foundations see

[7b] J. A. Wheeler, and W. H. Zurek, *Quantum Theory and Measurement*, Princeton University Press, Princeton, 1983.

[8] E. G. Beltrametti, and J. Lévy-Leblond (Eds.), *Advances in Quantum Phenomena*, NATO ASI Series B: Physics, Plenum, 1995.

[9] A. Aspect, and P. Grangier, *Experimental tests of Bell's inequalities*, 201–214, *ibid*.

[10a] A. Einstein, B. Podolsky, and N. Rosen, *Can quantum mechanical description of physical reality be considered complete*? Phys. Rev. 47(1935), 777-780.

See p. 142 in the Wheeler-Zurek volume [7]. Bohr's reply can be found in the same volume pp. 143–151 and was published in

[10b] N. Bohr, *Can quantum mechanical description of physical reality be considered complete*? Phys. Rev. 48(1935), 696–702.

For further insights into measurement theory and EPR experiments see

[10c] P. J. Lahti, and P. Mittelstaedt, (Eds.), *Symposium on the Foundations of Modern Physics*: 50 *years of the Einstein-Podolsky-Rosen Gedankenexperiment*, World Scientific, 1985 (look at the poems and songs!!).

[11a] S. Rozental (Ed.) *Niels Bohr*, Interscience, 1964, pp. 128–131. See also Bohr's article *Discussions with Einstein on epistemological problems in atomic physics*, in *Albert Einstein*: *Philosopher-Scientist*, P. A. Schilpp (Ed.), pp. 200–241, The library of living philosophers, Evanston, 1940. This is reproduced in

[11b] J. A. Wheeler, and W. H. Zurek, *Quantum Theory and Measurement*, Princeton University Press, Princeton, 1983, pp. 9–49.

[12a] R. Arens, and V. S. Varadarajan, *On the concept of Einstein-Podolsky-Rosen states and their structure*, J. Math. Phys., 41(2000), no. 2, 638–651.

[12b] D. Bohm, *Quantum Theory*, Prentice-Hall, 1951.

[13] P. A. M. Dirac, *The quantum theory of emission and absorption of radiation*, Proc. Royal Soc. London, Series A, 114(1927), 243–265.

[14] J. Schwinger (Ed.), *Quantum Electrodynamics*, Dover 1958.

[15] H. Weyl, *Theory of Groups and Quantum Mechanics*, Dover 1931.

[16] P. A. M. Dirac, *The Lagrangian in quantum mechanics*, Phys. Zeit. der Sowjetunion, 3(1933), 64–72.

[17] R. P. Feynman, *Space-time approach to non-relativistic quantum mechanics*, Rev. Modern Phys.,20(1948), 367–387.

[18] R. P. Feynman, *The concept of probability in quantum mechanics*, Proceedings of the Second Berkeley Symposium, J. Neyman (Ed.), University of California Press, Berkeley, 1951.

[19] B. Simon, *Functional Integration and Quantum Physics*, Academic Press, 1979.

[20] J. Glimm, and A. Jaffe, *Quantum Physics*: *A Functional Integral Point of View*, Springer, 1987.

[21a] D. Babbitt, *Wiener integral representations for certain semigroups which have infinitesimal generators with matrix coefficients*, Indiana Univ. Math. J. 19(1970), 1051–1067.

[21b] V. S. Varadarajan, and D. Weisbart, *Convergence of quantum systems on grids*, J. Math. Ana; Appl. 336(2007), 608–624.

[22a] M. Kac, *Enigmas of Chance*: *An Autobiography*, University of California Press, Berkeley and Los Angeles, 1987. Originally published by Harper and Row, 1985.

[22b] M. Kac, *On distributions of certain Wiener functionals*, in *Probability, Number Theory, and Statistical Physics. Selected Papers*, 268–280, 1979.

[22c] E. Nelson, *Feynman integrals and the Schrödinger equation*, Jour. Math. Phys. 5 (1964), 332–343.

Chapter 4
Super geometry*

Superalgebras serve as the basis for the construction of geometric objects, such as superprojective spaces and supermanifolds. The theory has applications to supergravitation in physics and it is studied by supermathematicians.
Igor Shafarevich

Supersymmetry was discovered by physicists around 1970. At first it was just a device for treating the bosonic and fermionc aspects of quantum statistical mechanics and quantum field theory on an equal footing. Eventually, due to the remarkable ideas and contributions of many physicists and mathematicians, supersymmetry came to be understood as the symmetry of a new kind of geometrical object, namely, a *super manifold*. The super manifolds are objects in *super geometry*, which is a deep generalization of conventional differential and algebraic geometry. The symmetries of super manifolds form *super Lie groups*.

4.1. The evolution of classical geometry
4.2. Super geometry
4.3. The theory of super manifolds
4.4. Super Lie groups

4.1 The evolution of classical geometry

It is common knowledge that the idea of symmetry and the role of group theory in describing it go back to very ancient times. Indeed, the remote origins of the subject had been discussed long ago in Hermann Weyl's beautiful book

* This essay and the next are based on various articles and lectures on supersymmetry I have given on many an occasion, especially a series of lectures in the Courant Institute in New York, in 2002, and in Oporto, Portugal, in 2006. I am extremely grateful to the large number of people with whom I have discussed the issues that form the core of these essays, and from whom I learnt most of the things I talk about. I would like to mention many notably Luigi Balduzzi, Claudio Carmeli, Gianni Cassinelli, Lauren Caston, Pierre Deligne, Sergio Ferrara, Rita Fioresi, Marian Lledo, Albert Schwarz, and Alessandro Toigo, to mention just a few. See also [1].

Symmetry. However it was only in the twentieth century that it became possible to erect a fully satisfactory mathematical theory of symmetry, namely the theory of Lie groups and their representations, that was adequate enough for all the applications. This theory, based on differential and algebraic geometry, was one of the major achievements of mathematics in the last century*. But, in the 1970's, the physicists, driven by their desire to explain the main features of the physical world at the most fundamental level, began to entertain the idea that the singularities in their descriptions of nature arose at least partly from the inadequacy of the usual models of spacetime to furnish the background for events taking place in regions of ultra-small dimensions, not to mention gravity where spacetime itself is part of what is to be described. Proceeding in this direction they came up with new models for spacetime and its symmetries. One of these is based on the conception that the description of spacetime should be based not only on the usual coordinates but on additional *Grassmann coordinates*, reflecting the fermionic structure of matter. This idea became the genesis of a new geometry, namely, *super geometry*. One has to thank the physicists for their daring and imagination to have come up with such a generalization. The effort to build a proper mathematical foundation for the discoveries of the physicists led to the development of an entirely new part of mathematics, namely the theory of super manifolds and super Lie groups. This is a development of geometry beyond its conventional limits and it is not clear at this time in what directions it will continue in the future. Certainly the happenings in the world of elementary particles and their fields will have a lot to say about the future evolution of geometry itself. I feel therefore that it is natural to start this essay with a brief overview of the evolution of the idea of space and geometry, starting with ancient times and leading gradually to superspace and super geometry.

From the historical point of view it is remarkable that already in 1854, Riemann, in his Göttingen inaugural address [2a, 2b, 2c], had speculated on the structure of space in its microscopic parts, and raised the possibility that the manifold structure may not be present in its infinitely small parts. He also advocated the point of view that we have to look for physics to tell us what this micro-structre should be. Here is what he wrote:

> Now it seems that the empirical notions on which the metric determinations of Space are based, the concept of a solid body and a light ray, lose their validity

* For a look at the development of Lie groups and their representations I refer the reader to a review I published in 2008 [2j]. It gives a bird's eye view of the historical development of Lie theory.

> in the infinitely small; it is therefore quite definitely conceivable that the metric relations of Space in the infinitely small do not conform to the hypotheses of geometry; and in fact, one ought to assume this as soon as it permits a simpler way of explaining phenomena.
>
> ..... An answer to these questions can be found only by starting from that conception of phenomena which has hitherto been approved by experience, for which Newton laid the foundation, and gradually modifying it under the compulsion of facts which cannot be explained by it. Investigations like the one just made, which begin from general concepts, can serve only to ensure that this work is not hindered by too restricted concepts, and that the progress in comprehending the connection of things is not obstructed by traditional prejudices.

Riemann's prophecy already started unfolding when Einstein discovered that *spacetime*, rather than space, is the invariant object attached to experience, and that its geometry is governed by a metric (in a pseudo Riemannian sense) of signature $(+\ -\ -\ -)$. If one thinks of all the various ideas about the micro-structure of spacetime that have come up in the meantime, one can only marvel at Riemann's intuition and understanding.

In order to put these developments in proper perspective we must begin by turning to the evolution of classical geometry which itself is tied intimately to the age-old problem of the structure of the space we live in. Indeed, for close to two thousand years, starting with Euclidean geometry, axiomatic geometry was thought to describe the geometry of physical space. Consequently attempts to build alternate geometries were viewed as heretical, forcing even such a titanic figure as Gauss to keep his investigations in noneuclidean geometry to himself, restricting himself to ironic comments in his letters to his scientific friends. However the discoveries of Bolyai and Lobachevsky, together with the posthumous publication of the work of Gauss, clearly showed that the problem of what type of geometry governs physical space is quite different from the problem of what types of geometry exist under different systems of axioms. The usual parallel axiom leads to euclidean geometry, while a different type of parallelism leads to the noneuclidean geometry of Bolyai and Lobachevsky. Nevertheless the intuition coming from our experience of the properties of the space we live in played a profound and mysterious role in determining alternative possibilities for what space can look like.

The modern era of geometry may be said to have started with Riemann's discovery of what we now call Riemannian geometry. One of Riemann's greatest contributions was the realization that *the geometry of a space must be built from its smallest parts*. These small parts can be described by *local coordinates* $x_1, x_2, \ldots, x_n$; and the infinitesimal distance, called the *metric*, between points with coordinates $(x_1, \ldots, x_n)$ and $(x_1 + dx_1, \ldots, x_n + dx_n)$

is given by a Pythagorean expression

$$\sum_{ij} g_{ij}(x_1, \ldots, x_n) dx_i dx_j$$

where Pythagorean means that the matrix (g_{ij}) is positive definite everywhere. Later, with the advent of Einstein's work and Minkowski's famous expositions of it, it was recognized that one can also have *pseudo Riemannian metrics* which are defined exactly as above but with the demand that the matrix (g_{ij}) have a fixed signature everywhere; if the signature is $(+ \; - \; \ldots \; -)$ the geometry is called *Minkowskian*. It became clear from Riemann's work that before one studies all the metrics that can be imposed on a space, one must first decide what sort of a space one has. In modern terms, one has to first set up the *differential topology* that the space ought to have, and then examine the possible metrics. For instance in studying compact Riemann surfaces, we classify the differential topology in terms of the *genus*, and then, for a given genus, study the complex structures that can be imposed on a surface of a given genus.

Riemann's lecture was published only after his death. It was a very difficult work to penetrate; for instance, it contained just a single formula. It introduced a cascade of revolutionary and profound ideas. It took almost a half century before Riemann's ideas were unraveled, by the efforts of differential geometers such as Bianchi, Christoffel, Levi-Civita, Ricci, and many others. This pioneering effort to understand Riemann was followed by the discoveries of Elie Cartan, and Hermann Weyl on global differential geometry and Lie theory. The view point of tensors dramatically changed the comprehension of natural laws, leading eventually to Einstein's great theory of universal gravitation as a manifestation of the curvature of a spaetime which is a global Minkowskian manifold of signature $(+ \; - \; - \; -)$.

Riemann's method of building spaces began to be be understood better when Hermann Weyl published his ground-breaking 1913 classic [2d]. It was continued by Hassler Whitney in his great paper [2e] on general differential manifolds which started a revolution in differential topology. The idea was to obtain the global space by *gluing together* small parts which can be completely described by *local coordinates*, together with a set of rules that tell us what *coordinate transformations* should *connect* these local coordinates on overlapping regions. This view emphasized the fact that many spaces would look the same locally, but would differ globally because of the

different possibilities in the gluing process. This approach allowed one to treat smooth, real and complex analytic manifolds in the same framework. It was using such a framework that Chevalley was able to give the first global treatment of Lie groups in his epoch-making book*. It was Chevalley's book that made the global view accessible to the general mathematical public [2f].

Then, in 1954, Serre [2g] showed how this method of constructing and studying manifolds by gluing local pieces could be extended to algebraic varieties defined over fields of *any characteristic*. This idea was generalized by Grothendieck and culminated in his theory of *schemes*, which apparently began as a series of remarks on Serre's paper. Grothendieck's ideas were developed in a monumental series of monographs published by the IHES, starting from [2h]. Serre's framework was already sufficient for Borel to begin the theory of algebraic groups over fields of any characteristic and then for Chevalley to develop the theory of semisimple Lie groups over arbitrary fields, culminating in his epoch-making classification of semisimple groups and their representations over arbitrary fields [2i].

Let us now return to the description of manifolds. In the first place coordinates are the results of measurements which are necessarily only over small regions because any measuring apparatus has only a small finite extent. In this discussion we must remember that in physical applications words like local and global have meanings depending on the context and must not be thought of in everyday terms. For instance local or small may refer to distances measured in magnitudes of the order of 10^{-16}cm in elementary particle experiments but the scale may change to millions of light years in describing the local structure of the universe. Moreover, although measuring originally referred to distances, eventually it came to be understood as a measurement of some aspects of the proximity properties of space that may be expressed numerically, or even measuring things like the pressure or temperature of some gas distributed throughout a star or a galaxy. Thus one replaces the idea of coordinates by the *ring* of all *numerical* quantities that are measured, the *local ring*. And space was to be thought of as a gluing together of these local rings. The mathematical tool to formulate this gluing is that of a *sheaf of rings of numerical functions*.

But then Hilbert and Gel'fand realized that for the most important types of local rings of functions, the points at which the functions of these rings

* Chevalley's book appeared in the famous Princeton Series with a dedication to Elie Cartan and Hermann Weyl. I still remember the blurb on the cover jacket saying that "the reader need no longer be afraid of shrinking neighborhoods of the identity element!"!

are defined can be obtained from the rings themselves by looking at all the maps of these rings into the number system we are working with. This is usually described by saying that the domain of the elements of the ring is the *spectrum* of the ring. For instance the homomorphisms of the polynomial ring $\mathbf{C}[x_1, \ldots, x_n]$ into $\mathbf{C}$ are precisely all the evaluation maps at points of $\mathbf{C}^n$ (Hilbert), while the homomorphisms of the ring $\mathbf{C}[X]$ of continuous functions on a compact Hausdorff space X are given by the evaluations at the points of X (Gel'fand). This made it clear that the function rings attached to the various regions of space already determine the local structure of space. The rings however have to be *mutually compatible*: when we take a ring associated to a region U and restrict the elements of it to an subregion V, we must get elements of the ring associated to V. This is expressed by saying that we have a *sheaf of rings*. The gluing must therefore match the sheaves on the individual parts that are being glued together.

There is another remarkable aspect of this point of view: if we have elements $f_1, \ldots, f_n$ of the ring attached to the region U, we can perform operations on these numerical functions and obtain new functions. Certainly, as the functions come from a ring, all algebraic operations are to be permitted. But in differential geometry where analysis is a great tool, we must allow arbitrary differentiable or analytic functions of $f_1, \ldots, f_n$. Thus by changing the scope of these operations we obtain the various kinds of geometry: algebraic, differential, analytic, holomorphic, and so on. By changing the number system we obtain real, complex, or p-adic manifolds. This is the modern perspective in geometry that has resulted in almost unlimited freedom in building models of space.

This perspective reached its climax with Grothendieck's theory of schemes; for a systematic exposition of the basics see [3a]. Proceeding by an intuition that was unlike that of anyone who preceded him, Grothendieck realized that a general theory that would be powerful enough to describe and solve all problems of algebraic geometry must be built on spaces where the local rings are *completely arbitrary commutative rings*. They would *not* be rings of numerical functions in general although they would still be formally attached to the regions of space and they would still be compatible; but now the restriction maps from a region to a subregion have to be part of the data defining the space. To recover some of the intuition lost in the process of making this generalization, the elements of the rings may be viewed as functions defined on the spectra of the rings. Thus the objects of study are topological spaces on

which we have a sheaf of commutative rings, the spaces themselves appearing as the spectra of these rings.

However, for an arbitrary commutative ring, the knowledge coming from representing elements of the ring as functions on the spectrum of the ring, is not complete because it loses information; there may be elements of the ring which give the value 0 at all points of the spectrum. For instance *nilpotent* elements have this property: indeed, if $x^n = 0$ and h is a homomorphism of the ring into a field, we have $h(x)^n = 0$ so that $h(x) = 0$. The presence of such elements in the local rings, as well as the desire to work with *all* number systems, make the theory of schemes technically very difficult to develop, but very rewarding once this is done, because it leads to the deepest results. The scheme thus became the most general type of space that can be created with the property that it can be built from its smallest parts which are spectra of arbitrary commutative rings. In the evolution of the concept of space the scheme thus represented the ultimate stage. It was the great achievement of Grothendieck and his successors that they were able to formulate, and, more often than not, solve, the fundamental questions of algebraic geometry, in the framework of schemes. Moreover, the language of schemes effected a surprising and profound unification of number theory with geometry.

Unbelievably, Grothendieck took the notion of space even farther. He realized from the general theory of schemes he had built that the underlying structure of a space was *completely determined by the category of sheaves defined on the scheme*. This led to his insight that the notion of space itself as consisting of points and their proximity properties (topology) is dispensable; one can replace it with the category whose objects represent the open sets and whose maps are the possible fundamental maps we allow in this category. The sheaves, viewed as functors from this category, form the *topos*. He then discovered the étale toposes of schemes that would enable him to connect the world of topology and differential geometry, the *continuous* world, with the world of number theory, the *discrete* world. I do not discuss this here since super toposes have not yet been used in super geometry. For a first glimpse of a topos see Allyn Jackson (p. 1052 of [3b]).

One can say that as long as the quest was to describe a *commutative world*, the notion of schemes and even toposes may be regarded as the end of the evolution. Riemann's prophetic warning against assuming that space is a manifold may thus be viewed as a cautionary reminder that even the schemes or toposes may be inadequate to describe the physical world as we acquire more powerful technical tools to probe into ever smaller regions of it. Already

quantum mechanics revealed that the algebras of physical quantities that come up in quantum theory are noncommutative. In supersymmetry we see the first indications that it may be worthwhile to relate these quantum algebras to spactime, and see if there are noncommutative aspects to spacetime itself.

4.2 Super geometry

To trace the beginnings of supersymmetry let us go to the one of the main features of quantum theory, namely, that elementary particles have an internal characteristic, their *spin*. The notion of spin, which is electromagnetic in its origin, allows one to make a crucial distinction between particles, namely, whether their spin is an integer or a half-integer. In the integer case they are called *bosons* while in the half-integer case they are called *fermions*. The distinction between bosons and fermions becomes crucial when one treats a *system* of identical particles. This may be called the *spin-statistics connection.* For an assembly of bosons there is no restriction as to whether they can all occupy the same state; indeed, the construction of lasers exploits this very possibility. Bosons are said to obey the *Bose-Einstein statistics*. On the other hand, the fermions obey Pauli's exclusion principle: no two fermions can occupy the same quantum state. As a result properties of matter differ dramatically in macroscopic phenomena depending on whether they are aggregates of bosons or fermions. The fact that fermions exclude each other has many consequences in phenomena of ultra low-temperature physics such as super fluidity and super conductivity. Fermions are said to obey the *Fermi-Dirac statistics*.

The spin-statistics connection is a deep consequence of fundamental results on the structure of quantum field theories. If $\mathcal{H}$ is the Hilbert space of a single particle, the Hilbert space of a system of N such particles is $S^N(\mathcal{H})$, the space of *symmetric* tensors of rank N, if the particle is a boson; and $\Lambda^N(\mathcal{H})$, the space of *skew symmetric* or *alternating* tensors of rank N, if the particle is a fermion. The equation

$$e \wedge e = 0,$$

so characteristic of exterior or Grassmann algebra, is the codification of the statement that two fermions cannot both be in the state defined by the same unit vector e. However in quantum field theory, it is possible that pairs of fermions give rise to bosons and vice versa; for instance an electron and positron may annihilate each other and produce radiation, or a photon may decompose into an electron–positron pair. The Hilbert space of quantum field

theory thus admits states in which there are particles of either type. Thus the Hilbert space of one particle states is an orthogonal direct sum $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$; here the vectors of $\mathcal{H}_0$ (resp. $\mathcal{H}_1$) define the states where the particle is a boson (resp. fermion). Supersymmetries then arise as transformations that exchange the bosonic and fermionic states.

In mathematical terms one may say that the Hilbert spaces of quantum field theory are $\mathbf{Z}_2$-graded. The attempt to have a unified way of treating bosons and fermions on the same footing soon led to linear algebra over $\mathbf{Z}_2$-graded, or, *super* vector spaces. A super vector space is a vector space V with a distinguished grading $V = V_0 \oplus V_1$ by the two element group $\{0, 1\} = \mathbf{Z}/2\mathbf{Z}$. V_0 and V_1 are called respectively the odd and even subspaces. They form a category admitting a tensor product, with specific rules, forming what is called a *tensor category*; I shall discuss this in a little more detail later. The result of this exercise is *super linear algebra.*

However the essential point of departure came when physicists, notably Salam and Strathdee [4a, 4b], realized that *one has to construct a graded generalization of classical geometry itself so that, after quantization, one obtains the graded Hilbert spaces so characteristic of quantum field theory in a natural manner.* Now the manifolds in classical geometry are studied using local coordinates. The new ideas then led to the creation of a generalization of classical geometry in which the usual local coordinates were supplemented by a set of *Grassmann coordinates*. The heuristic meaning of the Grassman coordinates was that they encoded the fermionic aspects of matter, in particular the Pauli exclusion principle that is at the basis of all properties of matter in the bulk. The automorphisms of such *super manifolds* were to be viewed as the supersymmetries, which form *super Lie groups*. The super manifolds, on quantization, led to the super Hilbert spaces and to the Fermi-Bose symmetries of quantum field theory. I must mention that the pioneering efforts of the (late) mathematician F. A. Berezin had an important role in the evolution of these ideas and in the creation of supersymmetry as a distinctive mathematical discipline.

With these works the following became clear: *the world itself is noncommutative, more precisely, supercommutative, and the quantum world of noncommutative algebras is a natural by-product of this super-commutative world*[*].

[*] The mathematician and physicist are going even further and looking at worlds which are even more noncommutative. I refer to the noncommutative geometry of people like Connes, Manin, Schwartz, Witten, Woronowicz, and others. I do not go into this terrain(!)

Because of the novelty and conceptual difficulty of this type of generalization of classical geometry, the notion of supersymmetry was formulated at first only *infinitesimally*. The infinitesimal supersymmetries form a *super Lie algebra*, which is a generalization, in the super linear world, of a Lie algebra. Indeed, the concept of a super Lie algebra itself was born in the attempts by physicists to formulate Fermi-Bose symmetry. Since a super Lie algebra is a linear object it was not difficult to formulate its definition precisely and analyze its properties. Gradually, the concept of supersymmetry was invested with a more fundamental meaning, and the idea took hold that supersymmetry was present in the early universe, only to be broken later when the universe started cooling. As a part of this evolution the goals of the physical side of the theory also became clear: to work out electrodynamics, and more generally, Yang-Mills theories, and Einstein gravity, under the umbrella of supersymmetry. For instance, Wess and Zumino worked out the supersymmetric version of electrodynamics, while Ferrara and Zumino treated nonableian gauge theories in the presence of supersymmetry. Also one now has a minimal supersymmetric standard model, and its predictions on the masses of elementary particles are in the list of things to be tested when the new collider at CERN starts operating. Finally, supersymmetry was enlarged to cover Einstein gravity by the ideas of Freedman, van Nieuwenhuizen, and Ferrara, as well as Deser and Zumino [6c].

The structure of these supersymmetric theories is such that they offer many advantages over conventional theories: softer divergences, reduction of possible Lagrangians, and so on. Although the global (as opposed to infinitesimal) formulation of supersymmetry was implicit in the treatments of the physicists, it was not fully revealed because of the difficulties in capturing the notion of super Lie groups (as opposed to the technically much simpler notion of super Lie algebras). In order to understand properly not only the infinitesimal but also the global geometric nature of super manifolds and supersymmetry at a fundamental level, it is therefore necessary to look more closely at the nature of super geometry as a generalization of classical differential and algebraic geometry. It turns out that what we need is a *generalization of Grothendieck's theory of schemes in which the local rings are supercommutative and have a special local structure, namely as a Grassmann algebra over commutative local rings*. The presence of the Grassmann variables, which are nilpotent and invisible, has the consequence that the mathematics of the theory of super manifolds can be developed cleanly only by using the methods and techniques of the theory of schemes. This method, called the *functorial*

method by mathematicians, is in fact extremely transparent, and is the one that mirrors completely all the calculations that the physicists make, so much so that one can have the advantage of working in the informal way that is most characteristic of physics, and yet lose almost nothing in mathematical rigor.

The Dirac mode

I begin with a quotation from Ferrara, one of the pioneers of the subject, from his introduction to [6c]:

> Supersymmetry, i.e., Fermi-Bose symmetry, is one of the most peculiar discoveries in the history of physics. The peculiarity lies in the fact that although a tremendous theoretical effort has been made in the field, no experimental evidence of this symmetry has shown up.

Part of the reason for the absence of experimental evidence for supersymmetry is that we may be very far from the energy levels needed for supersymmetry to be recreated. The collider being readied in CERN may be able to throw some light on this point.

The question whether theory should guide experiment or vice versa, has been a matter of intense discussion among physicists always. The greatest exponents of creating the theory first are of course Einstein and Dirac. Here is what the great physicist Y. Nambu calls the *Dirac mode* of doing physics. In particular Nambu points out that supersmmetry is an example of physics in the Dirac mode. First, here is Dirac himself, in his introduction to his famous paper on magnetic monopoles [5a]:

> The steady progress of physics requires for its theoretical formulation a mathematics that gets continually more advanced. This is only natural and to be expected... Noneuclidean geometry and noncommutative algebra, which were at one time considered to be purely fictions of the mind and pastimes for logical thinkers, have now been found to be very necessary for the description of general facts of the physical world. It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalisation of the axioms at the base of the mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation.
>
> The theoretical worker in the future will therefore have to proceed in a more indirect way. The most powerful method of advance that can be suggested at present is to employ all the resources of pure mathematics in attempts to perfect and generalise the mathematical formalism that forms the existing basis of theoretical physics, and *after* each success in this direction, to try to interpret the new mathematical features in terms of physical entities...

On the other hand, let us hear from Y. Nambu [5b]:

> The Dirac mode is to invent, so to speak, a new mathematical concept or framework first, and then try to find its relevance in the real world, with the expectation that (in a distorted paraphrasing of Dirac) a mathematically beautiful idea must have been adopted by God. Of course the question of what constitutes a beautiful and relevant idea is where physics begins to become an art.
>
> I think this second mode is unique to physics among the natural sciences, being most akin to the mode practiced by the mathematicians. Particle physics, in particular, has thrived on the interplay of these two modes. Among examples of this second approach, one may cite such concepts as
>
> Magnetic monopole
>
> Nonabelian gauge theory
>
> Supersymmetry
>
> On rare occasions, these two modes can become one and the same, as in the cases of Einstein gravity and the Dirac equation...

Some historical and scientific perspective

In addition to Salam and Strathdee there is a whole galaxy of physicists like Ferrara, Wess, Zumino, and the members of the Russian school such as Gol'fand, Likhtman, Volkov, Akulov, who contributed to the development of supersymmetry. On the mathematical side, first of all the late F. A. Berezin was one of the pioneers of super analysis and was very influential in the development of super geometry and super analysis as independent disciplines [6a, 6b]. The reader who wants to get a glimpse of the various historical and scientific strands should look at the monumental collection of articles (from the physicsts' point of view) edited by Ferrara [6c]. In addition I wish to mention a set of thought-provoking essays by Fronsdal, Flato, Hirai [6d], and by Freed [6e]. In addition the introduction of Grassmann algebras as number systems to do physics of fermions goes back to Schwinger's early papers [6f]. From the mathematicians' point of view there are many, starting with Deligne-Morgan [7a], Kostant [7b], Leites [7c], Manin [7d]; and more recently, the exposition by Varadarajan [8]. For a systematic treatment of super geometry as a generalization of both classical differential and algebraic geometry see the monograph by Carmeli, Caston, and Fioresi [9a]. For expositions from a different point of view see [9d, 9e, 9f].

4.3 The theory of super manifolds

The precise mathematical formulation of the concept of a super manifold with local commutative and Grassmann coordinates can now be given by adapting the definition of a classical manifold to the new situation through the notion of a ringed space, using the ideas of Grothendieck, as explained above.

The category of ringed spaces

If X is a classical smooth (C^∞ = infinitely differentiable) manifold, then for each open set $U \subset X$ we have the $\mathbf{R}$-algebra $C^\infty(U)$ of smooth real functions on U. The property of smoothness is *local* and the mathematical concept that captures this is that of a *sheaf*, expressed by saying that the assignment

$$U \longmapsto C^\infty(U)$$

is a *sheaf of function rings*. The rings $C^\infty(U)$ are often called the *local rings*. If X, Y are two smooth manifolds, a morphism, i.e., a smooth map, from X to Y is a continuous map ψ of X to Y such that for each open $U \subset Y$, the pull-back map ψ^* is a homomorphism of $C^\infty(U)$ into $C^\infty(\psi^{-1}(U))$. In this way we have the *category* of smooth manifolds and smooth maps. But in order to define supermanifolds where there are Grassmann coordinates locally that are numerically unobservable, the above notion of a sheaf of function rings has to be generalized significantly, the generalization being achieved in two stages.

The first step of such a generalization is already present in the notion, due to Grothendieck, of a *ringed space*. This is obtained by giving up the idea that the local rings consist of *numerical functions* defined on open sets U, but are rather *abstract commutative rings*, denoted usually by $\mathcal{O}_X(U)$. Unlike numerical functions the elements of these abstract local rings cannot be restricted to open subsets, and so we have to build the restriction maps as a part of the data defining a ringed space. Thus, for open sets U, V with $V \subset U$, we have restriction maps

$$\mathcal{O}_X(U) \longrightarrow \mathcal{O}_X(V), \qquad f \longmapsto f\big|_V$$

with the property that if $W \subset V \subset U$ are open, the restriction from U to W is the same as restricting first to V and then to W. The *local* nature of these

"functions" is expressed by the requirement that the assignment

$$U \longmapsto \mathcal{O}_X(U)$$

is a *sheaf of rings*. The sheaf axiom, which is the exact encoding of locality, is the following: if U is an open set and (U_i) is an open covering of U, and if $f_i \in \mathcal{O}_X(U_i)$ are given, then the following two statements are equivalent:

(a) there is a unique element $f \in \mathcal{O}_X(U)$ such that $f\big|_{U_i} = f_i$ for all indices i

(b) for any two indices i_1, i_2, $f_{i_1}\big|_{U_{i_1} \cap U_{i_2}} = f_{i_2}\big|_{U_{i_1} \cap U_{i_2}}$ (as elements of $\mathcal{O}_X(U_{i_1} \cap U_{i_2})$).

A *ringed space* is thus a pair $(X, \mathcal{O}_X = \mathcal{O})$ where X is a topological space and $\mathcal{O}$ is a sheaf of commutative rings on X. A *morphism* of a ringed space X into a ringed space Y is a continuous map $\psi : X \longrightarrow Y$ together with a pull-back map ψ^* of $\mathcal{O}_Y$ into $\mathcal{O}_X$ above ψ; this means that for each open $U \subset Y$, ψ^* is a homomorphism of $\mathcal{O}_Y(U)$ into $\mathcal{O}_X(\psi^{-1}(U))$. Notice that unlike the case when the rings of the sheaves are rings of functions, the pull-backs ψ^* for any open set $U \subset Y$ have to be specified as part of the data defining morphisms, and that these pull-backs have to be compatible with restrictions. One thus obtains the *category of ringed spaces and their morphisms*.

One may view the open sets in X as a category where the set of morphisms from U to V is empty unless $U \subset V$, in which case it consists of the natural inclusion of U in V. Then a presheaf is a contravariant functor from the category of open sets into the category of rings, and the morphisms of ringed spaces are natural transformations between these functors above maps between the spaces themselves.

In order to ascend to the concept of a super manifold, we have to take the second step of the generalization mentioned above which consists in replacing the commutative local rings by *supercommutative* local rings which have a local (not canonical) splitting as an algebra over commuting and anticommuting local coordinates. Supercommutativity means that the local rings are $\mathbf{Z}_2$-graded, and commutative in the super sense, that is, instead of the relation $ab = ba$ between arbitrary elements of the ring, we will have $ab = \pm ba$ where the sign is always plus unless both a and b are odd, in which case the minus sign is taken. Thus a super ringed space is a pair $(X, \mathcal{O}_X)$ where $\mathcal{O}_X$ is a sheaf of super commutative rings on the topological space X. A super manifold is then a super ringed space with the property that X is a classical

smooth manifold, and that locally the sheaf looks like the sheaf of rings on $\mathbf{R}^p$ of the form

$$U \longmapsto C^\infty(U)[\xi_1, \ldots, \xi_q] = C^\infty(U) \otimes \Lambda[\xi_1, \ldots, \xi_q]$$

where the ξ_j are Grassmann variables satisfying

$$\xi_i^2 = 0, \quad \xi_i \xi_j + \xi_j \xi_i = 0 \quad (i, j = 1, 2, \ldots, q)$$

and $\Lambda[\xi_1, \ldots, \xi_q]$ is the exterior algebra over $\mathbf{R}$ in q indeterminates. The reader who is familiar with quantum field theory will recognize in the requirement of super commutativity for the local rings the idea that to make the transition from a description of classical fields to fermi fields one should replace the classical commutators by anticommutators. An equivalent way of conceiving the idea of a super manifold is to start with the *superdomains* which are the ringed spaces $\mathbf{R}^{p|q} = (\mathbf{R}^p, \mathcal{O}^{p|q})$ where $\mathcal{O}^{p|q}$ is the sheaf $U \longmapsto C^\infty(U) \otimes \Lambda[\xi_1, \ldots, \xi_q]$; we then define a super manifold of dimension $p|q$ as the ringed space obtained by gluing together the $\mathbf{R}^{p|q}$.

It is clear that one can formulate the notion of a real, complex, or p-adic analytic, super manifold by replacing the underlying classical manifold by a real, complex, or p-adic analytic one. The natural end of this line of thought is the concept of a super scheme.

The category of super vector spaces and super algebras

As a basis for the study of super manifolds we have the theory of super vector spaces. If A is a ring and we have a direct sum decomposition $A = \oplus A_{i \in I}$ where the A_i are additive subgroups, we say that A is *graded by* I if I itself is an abelian group and $A_i A_j \subset A_{i+j}$ for $i, j \in I$. The A_i are sometimes referred to as the *homogeneous components* of A. The most common index group I is $\mathbf{Z}$, the group of integers. If V is a vector space and $S(V)$, $\Lambda(V)$ are the symmetric and exterior algebras of V, then $S(V)$ and $\Lambda(V)$ are graded in a natural manner by $\mathbf{Z}$. In supergeometry, grading by $\mathbf{Z}_2 = \mathbf{Z}/2\mathbf{Z}$ is fundamental; if A is $\mathbf{Z}_2$-graded, the nonzero elements of A_0 are called *even* and those of A_1 are called *odd*.

The infinitesimal structure of a supermanifold at a point is captured by the properties of its tangent spaces, which are objects of the category of *super vector spaces*, namely the category of $\mathbf{Z}_2$-graded vector spaces. This is a

category with direct sum $\oplus$, dual *, and tensor product $\otimes$, and is a prime example of a *tensor category* where the objects are abstract and the category has the three operations above with suitable functorial properties that imitate those of the vector space category. In particular we have isomorphisms c_{UV},

$$c_{UV} : U \otimes V \longrightarrow V \otimes U$$

which (through various requirements such as the hexagon axiom) capture the essential properties of the isomorphisms in the super vector space category. There the isomorphisms are given by

$$c_{UV} : u \otimes v \longmapsto (-1)^{p(u)p(v)} v \otimes u \quad (u \in U, v \in V).$$

Here $p(x)$ is the parity of x and is either 0 or 1. These structural assumptions explain all the characteristic features of the practical use of super vector spaces: sign rule, definition of supercommutativity, super Lie algebras, and so on. Moreover they "hide" all the signs in the various definitions. In what follows we shall discuss a few of these. The reader should note that almost everything makes sense for abstract tensor categories and this more general interpretation gives great insight [7a]. Finally the field over which all vector spaces, and more generally, the tensor category, is defined, is assumed to be of characteristic 0; in applications it is either $\mathbf{R}$ or $\mathbf{C}$.

For instance an algebra in the super category is an object A with a map $\mu : A \otimes A \longrightarrow A$. It is commutative (which means supercommutative) if $\mu = \mu \circ c_{AA}$. For super vector spaces this means the relation $ab = (-1)^{p(a)p(b)}ba$ between any two homogeneous elements a, b of A, a first example of the sign rule: *in any classical formula, whenever two odd elements are interchanged, there should appear a minus sign.* The c_{UV} can be generalized to an action of the symmetric group $\mathfrak{S}_N$ in N letters on $V \otimes V \otimes \cdots \otimes V$ (N factors). A Lie algebra in the tensor category is an algebra object $\mathfrak{g}$ with multiplication denoted by $[\cdot, \cdot]$, having the properties:

$$[\cdot, \cdot](1 + c_{\mathfrak{g}\mathfrak{g}}) = 0$$

$$[\cdot, [\cdot, \cdot]](1 + \sigma + \sigma^2) = 0$$

where σ is the automorphism of $\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}$ induced by the permutation $(123) \longrightarrow (312)$. If we unscramble these definitions we obtain the usual working definition of a super Lie algebra, as we shall see presently.

Historically, super Lie algebras first came up in the work of the physicists when the first supersymmetric infinitesimal transformations were written down. Gol'fand-Likhtman [10] and Volkov-Akulov [11] discovered the minimal SUSY extension of the Poincaré Lie algebra in the early 1970's. Wess-Zumino [12] discovered a little later, in 1974, the first example of a *simple* super Lie algebra, namely the minimal SUSY extension of the conformal Lie algebra. In 1975 V. Kac [13] formally defined super Lie algebras and carried out the super version of the Cartan-Killing classification of simple Lie algebras over $\mathbf{C}$.

Super Lie algebras

A *Lie super algebra* or *super Lie algebra* is a super vector space $\mathfrak{g}$ with a bracket $[\ ,\ ]$ which is a morphism from $\mathfrak{g} \otimes \mathfrak{g}$ to $\mathfrak{g}$ with the following properties:

(a) $[a, b] = -(-1)^{p(a)p(b)}[b, a]$.
(b) The (super) Jacobi identity (unscrambling the earlier version in a tensor category)

$$[a, [b, c]]+(-1)^{p(a)[p(b)+p(c)]}[b, [c, a]]+(-1)^{p(c)[p(a)+p(b)]}[c, [a, b]] = 0.$$

This is the working man's definition of the super Lie algebra. One can hide the signs above by rewriting these relations as we did previously, using $c_{\mathfrak{g}\mathfrak{g}}$ and the action of $\mathfrak{S}_3$ on $\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}$, showing that this is the same as saying that a super Lie algebra is simply a Lie algebra in the tensor category of super vector spaces. In fact in [7a] there is a beautiful treatment of Lie algebras in *any* k-linear tensor category where k is a field of characteristic 0. Clearly (b) reduces to the ordinary Jacobi identity for ordinary Lie algebras.

There is a second way to comprehend the notion of a super Lie algebra which is more practical. The bracket is skew symmetric if one of the elements is even and symmetric if both are odd. The super Jacobi identity has 8 special cases depending on the parities of the three elements a, b, c. By examining these we can conclude that a super Lie algebra is a super vector space $\mathfrak{g}$ on which a bilinear bracket $[\ ,\ ]$ is defined such that

(a) $\mathfrak{g}_0$ is an ordinary Lie algebra for $[\ ,\ ]$.
(b) $\mathfrak{g}_1$ is a $\mathfrak{g}_0$-module for the action $a \longmapsto \mathrm{ad}(a) : b \longmapsto [a, b]$ $(b \in \mathfrak{g}_1)$.
(c) $a \otimes b \longmapsto [a, b]$ is a symmetric $\mathfrak{g}_0$-module map from $\mathfrak{g}_1 \otimes \mathfrak{g}_1$ to $\mathfrak{g}_0$.

(d) For all $a \in \mathfrak{g}_1$, we have $[a, [a, a]] = 0$.

Except for (d) the other conditions are linear and can be understood within the framework of ordinary Lie algebras and their representations. The condition (d) is nonlinear and is the most difficult to verify in applications when Lie super algebras are constructed by putting together an ordinary Lie algebra and a module for it satisfying (a)-(c).

If A is a super algebra, we define

$$[a, b] = ab - (-1)^{p(a)p(b)}ba \qquad (a, b \in A).$$

It is then an easy verification that $[\ ,\ ]$ converts A into a super Lie algebra. It is denoted by A_L but often we omit the suffix L. If V is a super vector space, we write $\mathrm{End}(V)$ for the space of all linear maps $V \longrightarrow V$, regarded as a super vector space with the maps preserving the grading as even and the ones reversing the grading as odd. This is a super algebra and we often write $\mathfrak{gl}(V)$ for the corresponding Lie algebra; if $V = \mathbf{R}^{p|q}$ we write $\mathfrak{gl}(p|q)$ for $\mathfrak{gl}(V)$.

Let $\mathfrak{g}$ be a super Lie algebra and for $X \in \mathfrak{g}$ let us define

$$\mathrm{ad}\ X : \mathfrak{g} \longrightarrow \mathfrak{g}, \qquad \mathrm{ad}\ X(Y) = [X, Y].$$

Then

$$\mathrm{ad} : X \longmapsto \mathrm{ad}\ X$$

is a morphism of $\mathfrak{g}$ into $\mathfrak{gl}(\mathfrak{g})$. The super Jacobi identity is just the relation

$$[\mathrm{ad}\ X, \mathrm{ad}\ Y] = \mathrm{ad}\ [X, Y] \qquad (X, Y \in \mathfrak{g}).$$

The development of super linear algebra including the theory of modules over super commutative rings has no real surprises till we come to the generalization of the determinant, the *Berezinian*. Let R be a super commutative algebra over a field k of characteristic 0 and $R^{p|q}$ be the free module of dimension $p|q$ over R. Let $\mathrm{GL}(p|q)(R)$ be the group of invertible even morphisms of $R^{p|q}$. Then the **Berezinian** is a morphism of $\mathrm{GL}(p|q)(R)$ into $R_0^\times$ (the group of units of the even part R_0 of R) given by

$$\mathrm{Ber}(x) = \det(A - BD^{-1}C)\det(D)^{-1}$$

where

$$x = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

We have

$$\mathrm{Ber}(xy) = \mathrm{Ber}(x)\mathrm{Ber}(y).$$

This is the superversion of the determinant, discovered by F. A. Berezin. Since the entries of B and C are nilpotent, x is invertible if and only if A and D, whose entries are in the commutative ring R_0, are invertible, or even only invertible mod R_1. The usual determinant in the commutative category can be interpreted as the result of an action on the top part of the exterior algebra; for the super category the generalization of such an interpretation is homological and lies at a deeper level [7a].

Functor of points

In calculations with super manifolds physicists work with coordinates and manipulate the odd variables more or less on the same footing as the classical commuting coordinates. The functor of points approach is essentially a way to make sense of such calculations.

The philosophical-mathematical principle behind the functor of points approach to geometry that is revealed in Grothendieck's work is that an object can be comprehended by its relations with other objects [3b]. Put in this manner this seems to be a sort of well-meaning idea without substance. But translated into categorical terms and applied systematically it becomes extremely powerful. If $\mathcal{C}$ is a category and A is an object of $\mathcal{C}$, then the principle says that A is completely determined by the functor

$$F_A : T \longmapsto F_A(T) = A(T) = \mathrm{Hom}(T, A)$$

of $\mathcal{C}$ into sets, and that for two objects $A_1.A_2$ of $\mathcal{C}$, the set $\mathrm{Hom}(A_1, A_2)$ is in *natural bijection* with the set of *natural transformations* between the functors F_{A_1} and F_{A_2} (Yoneda's lemma). The elements of $F_A(T)$ are called the *T-points of A*.

This notion, applied to the category of super manifolds, leads to a concept of of points of a super manifold that is the true geometric one. Let X be a super manifold. For any super manifold T we write $X(T)$ for the set of maps $T \longrightarrow X$ and view the elements of $X(T)$ as the T-points of X. If T is a single point viewed as a purely even manifold, $X(T)$ is the set of

topological points of X. $X(T)$ is a contravariant functor in T and for any super manifold Y, the set $\mathrm{Hom}(X, Y)$ can be recovered precisely as the set of natural transformations $X(T) \longrightarrow Y(T)$ (Yoneda's lemma). One can use this to define the products of super manifolds: the product $X_1 \times \cdots \times X_N$ is the super manifold whose T-points are $X_1(T) \times \cdots \times X_N(T)$. In this context we introduce the usual definition of when a set-valued contravariant functor $T \longmapsto f(T)$ is *representable*, namely when there is a super manifold M such that $M(T) = f(T)$.

Morphisms

Because the local rings $C^\infty(x_1, \ldots, x_p)[\theta_1, \ldots \theta_q]$ are not *algebraically* generated by the $x_1, \ldots, x_p, \theta_1, \ldots \theta_q$, one has to be careful in defining morphisms. Actually the situation is exactly as in the algebraic case, and this is the reason why the theory of super manifolds can be developed for the most part as the theory of schemes. If M, N are super manifolds and $x_1, \ldots x_p, \theta_1 \ldots \theta_q$ are coordinates on N, to define a morphism from M to N it is sufficient to specify the images of the x_i, θ_j. To illustrate this I give a characteristic example of a supersymmetic transformation that mixes the odd and even variables. The manifold is $\mathbf{R}^{1|2}$ with coordinates t, θ_1, θ_2. The transformation is then specified by $t' = t + \theta_1\theta_2, \theta_1' = \theta_1, \theta_2' = \theta_2$. If $g \in C^\infty(U)$ where $U \subset \mathbf{R}$ is open, the image of g is formally $g(t + \theta_1\theta_2)$; if g is a polynomial in t this is obviously well-defined; if g is smooth but not a polynomial, we give a meaning to $g(t + \theta_1\theta_2)$ by expanding it as a Taylor series. The Taylor series terminates since $(\theta_1\theta_2)^2 = 0$. Thus

$$g(t + \theta_1\theta_2) = g(t) + g'(t)\theta_1\theta_2.$$

This method is perfectly general.

Differential calculus

Because of the above result on morphisms, one has a theory of differentiation on super manifolds that is completely analogous to the theory on classical manifolds. The differential criteria for a map to be a local diffeomorphism, the structure of submanifolds, etc remain essentially the same. Super submanifolds of super manifolds are defined by local equations whose differentials

have maximal rank at each point. For instance if

$$f_1, \ldots, f_r, g_1, \ldots, g_s$$

are r even and s odd sections of $\mathbf{R}^{p|q}$ such that the matrices

$$\left(\frac{\partial f_a}{\partial x^i}\right), \qquad \left(\frac{\partial g_b}{\partial x \theta^j}\right)$$

have respective ranks r, s at 0, then the equations

$$f_1 = \cdots = f_r = g_1 = \cdots = g_s = 0$$

define a sub super manifold locally at 0 of dimension $(p-r)|(q-s)$.

Integral calculus

Let

$$\theta^I = \theta^{i_1}\theta^{i_2}\ldots\theta^{i_k} \qquad I = (i_\mu), \{i_1 < i_2 < \cdots < i_k\}.$$

On

$$\Lambda = \mathbf{R}[\theta^1, \ldots, \theta^q]$$

the integral is a linear map

$$a \longmapsto \int a d^q\theta$$

defined by

$$\int \theta^I d^q\theta = \begin{cases} 0 & if |I|<q \\ 1 & if I=:Q=\{1,2,\ldots,q\}. \end{cases}$$

Integration is also differentiation:

$$\int = \left(\frac{\partial}{\partial\theta^q}\right)\left(\frac{\partial}{\partial\theta^{q-1}}\right)\cdots\left(\frac{\partial}{\partial\theta^1}\right).$$

In the local ring with coordinates x^i, θ^j,

$$\int s d^p x d^q\theta = \int s_Q d^p x \qquad (s = \sum_I s_I \theta^I).$$

The change of variables formula

For a morphism given locally as

$$\psi : (x, \theta) \longmapsto (y, \varphi)$$

we define the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial y}{\partial x} & -\frac{\partial y}{\partial \theta} \\ \frac{\partial \varphi}{\partial x} & \frac{\partial \varphi}{\partial \theta} \end{pmatrix}.$$

Then

$$\int s = \int \psi^*(s)\mathrm{Ber}(J\psi)$$

for compactly supported sections of the local ring. For arbitrary manifolds we use partitions of unity as in the classical case. This beautiful formula goes back to Berezin. The justification for the peculiar definition of integration in the anticommuting variables is the change of variables formula.

The global sections functor

To any C^∞ super manifold M we have the super algebra of global sections $\mathcal{O}_M(M) = A(M)$. It is possible to show that M can be recovered from $A(M)$. More precisely,

$$M \longmapsto A(M)$$

is a functor from the category of smooth super manifolds to the category of super algebras which is fully faithful[7b].

Heuristic conception of a super manifold

The intuitive picture of M is that of a *classical smooth manifold* $|M|$ *surrounded by a Grassmannian cloud [6e, 8]*. The cloud cannot be seen: in any measurement the odd variables will be 0 because they are nilpotent. Thus measurement sees only the underlying classical manifold $|M|$. Nevertheless the presence of the cloud eventually has consequences that are striking.

As an illustration of how we use the idea of functor of points even in elementary contexts almost without thinking, consider the polynomial ring $F[t]$ where F is a field and t is an *indeterminate*. One can perform calculations

in this ring purely abstractly, but somehow when we first encounter it the elements of the ring appear elusive. But suppose now we take a ring G containing F; any polynomial $f \in F[t]$ can now be evaluated at a point of G and so f gives rise to a function from G to G. By varying G we can get a complete hold on f through very concrete numerical functions. Suppose now we consider the superdomain $\mathbf{R}^{1|1}$. The abstract way is to operate on the sheaf with elements of the form $f + \theta g$ where θ is the Grassmann variable and f, g are smooth functions of the real coordinate t. But if T is any super manifold and we consider a map of T into $\mathbf{R}^{1|1}$, the pull-back of such a map will map t into an even section on T and θ into an odd section on T. We may thus interpret t and θ as even and odd sections of T so that $\mathbf{R}^{1|1}(T)$ becomes identified with the *set* of all (t, θ). This method of working with the functor of points allows us to calculate in a way that is very close to the physicists' way of using θ as a Grassmann indeterminate.

For a super manifold A, the functor $T \longmapsto A(T)$ is huge in the sense that its domain of definition is the category of *all* super manifolds. For practical purposes one may want functors with more limited domains that may still allow us to recover A. One candidate is the category $\mathcal{E}$ of all finite dimensional exterior algebras over $\mathbf{R}$. If Λ is one such we can think of Λ as the structure sheaf of a super manifold whose classical reduction is just a single point; we may call this a *super point* denoted by p_Λ. Thus we can restrict the functor of points to the category of super points and obtain the functor

$$E_A : \Lambda \longmapsto A_\Lambda = \mathrm{Hom}(p_\Lambda, A) \approx \mathrm{Hom}(\mathcal{O}_A(A), \Lambda)$$

from the category $\mathcal{E}$ to the category of sets. However, because the domain of this functor is very small, it is not possible to recapture the super manifold A from the functor E_A as long as we think of it as a functor whose values are just sets. But it turns out that the sets A_Λ acquire the structure of a Λ_0-modules (Λ_0 is the even part of Λ), and $\mathrm{Hom}(A_1, A_2)$ is in natural bijection with the natural transformations between the functors E_{A_1} and E_{A_2} preserving this manifold structure. This follows from the work of Schwarz and Voronov [9b].

Berezin's work already had the idea of considering only the exterior algebra. Actually this idea goes back to A. Weil in classical differential geometry. In a beautiful paper (*Théorie des points proches sur les variétés différentiable*, Colloque de Géometrie Differentielle(Strasbourg 1953), C. N. R. S.,

111–117; also in *Andre Weil:Oeuvres Scientifiques:Collected Papers*, Vol II, 103–109, Springer 1979). Weil introduces commutative local algebras of finite dimension A and considers the A-points of a smooth manifold to be the maps from the smooth functions on the manifold to A. This can be extended to the super context by replacing these algebras by supercommutative local algebras and has been done in [9c]. The corresponding functors are called the *Weil-Berezin functors*.

Representability of the various functors is an important issue. But even when these functors are not representable, there is some advantage in thinking of them as *virtual manifolds* and the natural maps between them as *virtual morphisms*.

4.4 Super Lie groups

Super Lie groups act on super manifolds and so provide the symmetries of these new objects. They can be defined categorically as group objects in the category of super manifolds. Thus, to say that G is a super Lie group, is to say that G is first of all a super manifold, and that we have morphisms

$$\mu : G \times G \longrightarrow G \text{ (multiplication)} \quad \iota : G \longrightarrow G \text{ (inverse)}$$

that have the associativity properties usually ascribed to multiplication and inverse, together with the identity element

$$e : \{\text{point}\} \longrightarrow G.$$

From the perspective of functor of points the functor $G(T)$ takes values in groups, so that a super Lie group may be defined as a contravariant functor from the category of super manifolds to the category of groups, which is representable. The theory of super Lie groups may be built up in the same way as the theory of classical Lie groups, but some of the technical aspects become more involved because things have to be done using the functor of points.

One can define on any super Lie group the notion of left (or right) invariant vector fields, and hence the notion of a Lie algebra of a super Lie group. This is a super Lie algebra and the correspondence between G and its super Lie algebra $\mathfrak{g}$ has essentially all the features of the classical correspondence.

The simplest super Lie group of some generality is $\mathrm{GL}(p|q)$. I have nor specified the ground field which can be $\mathbf{R}$, $\mathbf{C}$ or a nonarchimedean local field.

Let us assume for simplicity that it is $\mathbf{R}$. We can think of this as the set of $p|q$ block matrices

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where the entries of A and D are even variables, and the entries of B and C are odd Grassmann variables, with A and D invertible. This means that the underlying classical Lie group is $G_0 = \mathrm{GL}(p) \times \mathrm{GL}(q)$ with coordinates as the entries of the matrices A and D, and the sheaf is

$$\mathcal{O}_{G_0}[B, C].$$

If T is a super manifold and we interpret the entries of $A(T)$ and $D(T)$ as even sections on T and those of $B(T)$ and $C(T)$ as odd sections, then the functor of points is

$$T \longmapsto G(T)$$

where $G(T)$ is the group of elements of the matrix

$$\begin{pmatrix} A(T) & B(T) \\ C(T) & D(T) \end{pmatrix}$$

with $A(T)$, $D(T)$ invertible.

Other super Lie groups are, as usual, defined as sub super Lie groups of $\mathrm{GL}(p|q)$. $\mathrm{SL}(p|q)$: This may be defined as the kernel of the morphism

$$\mathrm{Ber} : \mathrm{GL}(p|q) \longrightarrow \mathrm{GL}(1|0)$$

where Ber is the Berezinian, or, equivalently, as the subgroup defined by the equation

$$\mathrm{Ber} = 1.$$

Similarly, but in a more involved manner, we can define the orthosymplectic groups $\mathrm{OS}(p|2q)$. See [8], pp. 290–294.

It is possible to discuss these groups as examples of *super algebraic groups*. For a construction of the *Chevalley super groups* see [14].

Super Harish-Chandra pairs

Of great importance for many applications is a theorem which allows one to describe a super Lie group in more concrete terms. Given a super Lie group

G we have a pair $(G_0, \mathfrak{g})$ where G_0 is the classical Lie group underlying G, and $\mathfrak{g}$ is the super Lie algebra of G. We have

(1) $\mathrm{Lie}(G_0) = \mathfrak{g}_0$.
(2) G_0 has an action on $\mathfrak{g}$ and its differential is the adjoint action of $\mathfrak{g}_0$ on $\mathfrak{g}$.

Such pairs $(G_0, \mathfrak{g})$ can be also introduced *a priori* without reference to a super Lie group. They are then called *super Harish-Chandra pairs*. The basic theorem, which is fundamental for us, is the result that

$$G \longmapsto (G_0, \mathfrak{g})$$

is a covariant functor which is an equivalence of categories from the category of super Lie groups to the category of super Harish-Chandra pairs. In terms of super Harish-Chandra pairs one can describe actions of super Lie groups on super manifolds. This allows a satisfactory treatment of super homogeneous spaces $X = G/H$ where G is a super Lie group and H is a closed super Lie subgroup. The functor of points of X is *not* the functor $T \longmapsto G(T)/H(T)$ but rather the *sheafification* of it [15, 16a, 16b]. For a concise and beautiful presentation of the basic theory of super Lie groups, see [17].

References

[1] V. S. Varadarajan, *Unitary representations of super Lie groups and some applications*, The Math. Student Special Centenary Volume, (2007), 33–57.

[2a] Riemann's inaugural address is included in his Collected Works:

Bernhard Riemann: Collected Papers, R. Narasimhan, (ed.), Springer Verlag, 1990.

There are English translations of Riemann's talk by W. K. Clifford and M. Spivak. In addition, the Collected Papers contains Weyl's annotated remarks on Riemann's talk.

[2b] W. K. Clifford, *On the space-theory of matter*, Mathematical Papers, Chelsea, 1968.

[2c] M. Spivak, *Differential geometry*, Vol II, Publish or Perish Inc, 1979.

[2d] H. Weyl, *The Concept of a Riemann Surface*, Dover, 2009. This is a reproduction of the translation of the second edition of the classic original which was published by B. G. Teubner in 1913 and was called *Die Idee der Riemannschen Fläche*, based on his lectures at Göttingen in the years 1911–1912.

[2e] H. Whitney, *Differentiable manifolds*, Ann. of Math. (2) 37 (1936), no. 3, 645–680.

[2f] C. Chevalley, *Theory of Lie groups. I.* Princeton University Press, Princeton, N. J., 1946, 1957. The 1946 edition was the first global treatment of Lie groups. It instantly transformed the subject, giving access to one of the central areas of mathematics.

[2g] J. -P. Serre, *Faisceaux algébriques cohérents*, Ann. of Math. (2) 61, (1955). 197–278. See also *Oeuvres*, Vol I., 310–391, Springer, 2003 (Paperback edition).

[2h] A. Grothendieck, *Éléments de géométrie algébrique. I. Le langage des schémas.*, Inst. Hautes Études Sci. Publ. Math. No. 4 (1960).

[2i] C. Chevalley, *Classification des groupes algébriques semi-simples.* [*The classification of semisimple algebraic groups*]. Collected works. Vol. 3. Edited and with a preface by P. Cartier. With the collaboration of Cartier, A. Grothendieck and M. Lazard. Springer-Verlag, Berlin, 2005.

[2j] V. S. Varadarajan, *Lie Groups. An approach through invariants and representations.* Review of a book by C. Processi. Bull. Amer. Math. Soc., 45 (2008), 661–674.

Erratum, Bull. Amer. Math. Soc., 46(2009) 175–175.

[3a] R. Hartshorne, *Algebraic Geometry*, Springer, 1977.

For a beautiful study of Grothendieck see Allyn Jackson's essays:

[3b] A. Jackson, *Comme appelé du néant—as if summoned from the void: the life of Alexandre Grothendieck. I, II.*, Notices of the Amer. Math. Soc. 51 (2004), 1038–1056; 1196–1212.

[4a] A. Salam, and J. Strathdee, *Super-gauge transformations*, Nuclear Physics B76 (1974) 477–482. See the collection [6c], Vol 1. pp. 35–40, below.

[4b] A. Salam, and J. Strathdee, *Superfields and Fermi-Bose symmetry*, Phys. Rev. D., 11(1975) 1521–1535.

Salam and Strathdee pioneered the approach where supermanifolds are introduced and superfields are just sections of the structure sheaf of these supermanifolds.

[5a] P. A. M. Dirac, *Quantized singularities in the electromagnetic field*, Proc. Roy. Soc. Lond., A133(1931), 60–72.

[5b] Y. Nambu, *Broken Symmetry: Selected Papers of Y. Nambu*, 1995, World Scientific, pp. 383–389, eds. T. Eguchi and K. Nishijima.

[6a] F. A. Berezin, *Introduction to Superanalysis*, D. Reidel, 1987. Edited with a foreword by A. A. Kirillov and an appendix by V. I. Ogievstsky. Translated from Russian by J. Niederle and R. Kotecký. Translation edited by D. A. Leites.

For an account of Berezin's work and papers related to his interests see the following.

[6b] R. L. Dobrushin, R. A. Minlos, M. A. Shubin, and A. M. Vershik, (Eds.), *Contemporary Mathematical Physics: F. A. Berezin Memorial Volume*, AMS Translations, Ser. 2, Vol. 175, 177, 1996.

For a comprehensive collection of the fundamental papers pertaining to supersymmetery see the following.

[6c] S. Ferrara, (Ed.), *Supersymmetry*, vols. 1, 2. World Scientific, Singapore, 1987.

[6d] C. Fronsdal (Ed.), *Essays on Supersymmetry, with contributions by* M. Flato, C. Fronsdal, and T. Hirai, Mathematical Physics Studies, Vol. 8, 1986, D. Reidel.

[6e] D. S. Freed, *Five Lectures on Supersymmetry*, Amer. Math. Soc., R. I., 1999.

[6f] J. Schwinger, *Exterior algebra and the action principle*, Proc. Nat. Acad. Sci. USA., 48(1963), 603–611. Also in *Selected Papers* (1937–1976) *of Julian Schwinger*, M. Flato, C. Fronsdal, K. A. Milton (Eds.), D. Reidel, 1979.

[7a] P. Deligne, and J. Morgan, *Notes on supersymmetry* (*following Joseph Bernstein*), in *Quantum Fields and Strings: a Course for Mathematicians*, vol. I, 41–97, American Mathematical Society, Providence, R.I., 1999.

[7b] B. Kostant, *Graded manifolds, graded Lie theory, and prequantization* in *Differential Geometrical Methods in Mathematical Physics*, Proc. Symp. Univ. Bonn, Bonn, 1975, Lecture Notes in Mathematics, Springer, Berlin, 1977.

[7c] D. Leites, *Introduction to the theory of super manifolds*, Usp. Mat. Nauk. 35(1), 3-57, 1980, Russ. Math. Surveys, 35(1), 1–64, 1980.

[7d] Yu. I. Manin, *Gauge Field Theory and Complex Geometry*, Grundlehren der Mathematischen Wissenschaften, 289, Springer, Berlin, 1988.

[8] V. S. Varadarajan, *Supersymmetry for Mathematicians: An Introduction*, AMS-Courant Institute Lecture Notes, Providence, R.I., 2004.

[9a] C. Carmeli, L. Caston, and R. Fioresi, (with an appendix by I. Dimitrov) *Mathematical Foundations of Supersymmetry*, preprint, 2008.

[9b] A. S. Schwarz, *On the definition of superspace*, Teor. Mat. Fiz., 60(1984), 37–42.

A. A. Voronov, *Maps of supermanifolds*, Teor. Mat. Fiz., 60(1984), 43–48.

[9c] L. Balduzzi, C. Carmeli, and R. Fioresi, *The local functors of points of supermanifolds*, to appear in Expositiones Mathematicae.

This paper discusses the ideas of Weil, Berezin, Schwartz and Voronov in the general context of finite-dimensional super Weil algebras.

[9d] M. Batchelor, *Two approaches to supermanifolds*, Trans. Amer. math. Soc., 258(1980), 257–270.

[9e] C. Bartocci, U. Bruzzo, and D. H. Ruipérez, *The Geometry of Supermanifolds*, in the series *Mathematics and its Applications*, Kluwer, 1991.

[9f] B. DeWitt, *Supermanifolds*, Cambridge monographs on mathematical physics, Cambridge University Press, Cambridge, 1984.

[10] Yu. A. Gol'fand and A. P. Likhtman, *Extension of the algebra of the Poincaré group generators and violation of P invariance*, JETP Lett., 13(1971), 323–326.

[11] D. V. Volkov, and V. P. Akulov, *Is the neutrino a Goldstone particle*? Phys. Lett. 46B(1973), 109–110.

[12] J. Wess and B. Zumino, *Supergauge transformations in four dimensions*, Nucl. Phys., B70(1974), 39–50.

[13] V. G. Kac, *Lie superalgebras*, Adv. Math., 26(1977), 8–96.

[14] R. Fioresi, and F. Gavarini, *Chevalley supergroups*, arXiv:0808.0785v2 [math.RA], 2008.

[15] R. Fioresi, M. A. Lledó, and V. S. Varadarajan. *The Minkowski and conformal superspaces*, J. Math. Phys., 48(11)113505, 27, 2007.

[16a] L. Balduzzi, C. Carmeli, and G. Cassinelli, *Super G-spaces*, 159–176, in *Symmetry in Mathematics and Physics*, Contemp. Math., 490, Amer. Math. Soc., 2009.

[16b] L. Balduzzi, C. Carmeli, and R. Fioresi, *Quotients in super geometry*, 177–187, in *Symmetry in Mathematics and Physics*, Contemp. Math., 490, Amer. Math. Soc., 2009.

[17] C. Carmeli, *Super Lie groups : Structure and Representations*, Thesis, University of Genoa, 2006.

Chapter 5
Unitary representations of super Lie groups*

> There is still no direct evidence that supersymmetry is a symmetry of the physical world, that elementary particles arrange themselves in supermultiplets of spin differing by half a unit. It must be broken, since if unbroken it would predict that the particles in a supermultiplet have the same mass. Finding the correct breaking mechanism is probably still the basic unsolved problem.
> *Bruno Zumino*

The concept of a unitary representation(UR) of a super Lie group is formulated via super Harish-Chandra pairs. For *super semidirect products* the classical Wigner-Mackey theory of little groups works perfectly in the supersymmetric setting, and leads to a description of all of their unitary irreducible representations (UIR). The Clifford structure of the representations and the concept of super multiplets all make sense in the general context of super semidirect products, which includes all cases studied by the physicists, and leads to many of their major predictions: multiplet structure (both for minimal and extended supersymmetry), and the famous *susy partners*.

5.1. Unitary representations of a super Lie group
5.2. Super imprimitivity theorem for even homogeneous spaces
5.3. Unitary irreducible representations of super semidirect products
5.4. Super Poincaré Lie algebras and Lie groups
5.5. Unitary representations of super Poincaré groups
5.6. Super particles and their multiplet structure

5.1 Unitary representations of a super Lie group

In quantum physics symmetries are expressed in terms of unitary operators. If $P(\mathcal{H})$ is the projective space of the Hilbert space $\mathcal{H}$ of quantum states, and we think of a symmetry as a bijection of $P(\mathcal{H})$ that preserves the transition

* This essay is partly based on the lectures I gave at a conference on supersymmetry held in Oporto, Portugal, in 2006. I am grateful to the organizers of that conference for their splendid hospitality. I also wish to thank Claudio Carmeli, Gianni Cassinelli, Dan Freed, Marian Lledo, and Albert Schwarz for several stimulating conversations during the conference. See also [1a, 1b].

probabilities $|(\phi, \psi)|^2$, then it is induced by a unitary or anti-unitary operator of $\mathcal{H}$, unique upto a scalar of absolute value 1 (phase factor): if L is the operator, $[\psi] \longmapsto [L\psi]$ is the induced symmetry, $[\psi]$ being the point in $P(\mathcal{H})$ corresponding to the vector ψ. If we want symmetry with respect to a topological group G, then on a closed normal subgroup G_0 of G of index at most 2 all symmetries are unitary; in particular, $G = G_0$ if G is connected. The phase factors may present problems and one may have to go to a central extension of the group G to get unitary representations without any phase factors. In any case symmetry is ultimately expressed via the notion of a unitary representation.

To formulate the notion of unitary representations in the supersymmetric context it is natural to proceed first *infinitesimally*. The study of *unitary representations of super Lie algebras* began not long after super Lie algebras were discovered. The physicists, proceeding infinitesimally, unveiled a rather complete picture of the elementary particles and their states in the super world, based on the heuristic assumption that the transition to the super world would not affect the main outlines of the techniques (little groups) used to determine the structure of elementary particles in the usual setting.

However unitary representations of *super Lie groups* are a different story because of the subtle manner in which a super Lie group differs from an ordinary Lie group. A super Lie group is an object which is not a group but a group-valued functor, and it is not immediately clear how to mesh the functorial approach with the functional analytic aspects that enter when dealing with representations that may be *infinite dimensional*. Extra care is therefore needed to define a unitary representation of a super Lie group in a possibly infinite dimensional super Hilbert space. The study of unitary representations of a *super Lie group* therefore lagged far behind that of the representations of super Lie algebras.

In this section I shall discuss how to make the correct definition for a unitary representation of a super Lie group. It seems likely that further developments of the representation theory of super Lie groups may require a theory of representations of super Lie groups in Banach and even Frechet spaces, but we shall not take up these generalizations here. It is however very important to be completely general and not restrict the class of super Lie groups in any way, in defining the fundamental concepts.

Our point of departure is the fact that a super Lie group may be identified with a super Harish-Chandra pair $(G_0, \mathfrak{g})$. This is reasonable, since as I have mentioned in the previous essay, the category of super Lie groups is naturally

equivalent to the category of super Harish-Chandra pairs. Let me recall that $(G_0, \mathfrak{g})$ is a super Harish-Chandra pair if G_0 is a classical Lie group and $\mathfrak{g}$ is a super Lie algebra with an action of G_0 on it such that

(i) $\mathrm{Lie}(G_0) = \mathfrak{g}_0$=the even part of $\mathfrak{g}$.
(ii) The action of G_0 on $\mathfrak{g}$ is the adjoint action of G_0; more precisely, the adjoint action of $\mathfrak{g}_0$ on $\mathfrak{g}$ is the differential of the action of G_0 on $\mathfrak{g}$.

Thus we need not make any distinction between super Lie groups and super Harish-Chandra pairs. If we carry over this principle to representations we are led to the idea that a representation of a super Lie group $(G_0, \mathfrak{g})$ may be thought of as a triple $(\pi, \gamma, \mathcal{H})$ where π is an *even* representation of G_0 in a super Banach (or Frechet) space $\mathcal{H}$ (this means that $\pi(g)$ is even for all $g \in G_0$), and γ is a super representation of $\mathfrak{g}$ in $\mathcal{H}$ compatible with π. We shall only be interested in the case of *unitary* representations. This means that $\mathcal{H}$ is a super Hilbert space, π_0 is unitary, and γ satisfies an appropriate condition that reflects the unitarity of the representation at the infinitesimal level. It turns out that a natural condition would be to require that $\gamma(X)$ be skew self adjoint in the super sense for all $X \in \mathfrak{g}$. There is in addition an important point, technical, but critical nevertheless. To see this, we note that for $X \in \mathfrak{g}_1$, $[X, X] \in \mathfrak{g}_0$ and $\gamma([X, X]) = 2\gamma(X)^2$. Now, the operators corresponding to the elements of $\mathfrak{g}_0$ are typically unbounded, and so the operators $\gamma(X)(X \in \mathfrak{g}_1)$ will in general be unbounded; this means that there are domain considerations to be taken care of. In the next paragraph I shall address both of these points.

The basic definitions

All sesquilinear forms are linear in the first argument and conjugate linear in the second (opposite to the physicists' convention). A *super Hilbert space* is a $\mathbf{Z}_2$-graded Hilbert space $\mathcal{H}$ where the $\mathcal{H}_i (i = 0, 1)$ are closed mutually orthogonal subspaces. If $\langle x, y\rangle$ is defined as $i^{p(x)p(y)}(x, y)$, then $\langle x, y\rangle$ is an even super Hermitian form. For bounded operators $T : \mathcal{H} \longrightarrow \mathcal{H}$ the adjoint with respect to the super form is denoted by $T^\dagger$ and is related to the usual Hilbert space adjoint by $T^\dagger = T^*$ or $-iT^*$ according as T is even or odd. We define $T^\dagger$ for unbounded T by this rule in terms of T^*.

In order to motivate the definition of unitary representations of super Lie groups we begin by observing that in the even case of an ordinary Lie group,

the unitarity of a representation π of the group is usually expressed at the infinitesimal level as follows: if $d\pi$ is the representation of the Lie algebra, then for all X in the Lie algebra the operators $d\pi(X)$ are skew Hermitian. If we carry this over to the super setting the condition should be that the $d\pi(X)$ should be skew super Hermitian for all $X \in \mathfrak{g}$. This means that for $X \in \mathfrak{g}_0$ we should require that $d\pi(X)$ should be skew Hermitian while for $X \in \mathfrak{g}_1$ we should have $d\pi(X)^* = -id\pi(X)$, ignoring domain considerations. Let

$$\zeta = e^{-i\pi/4}.$$

Then this last condition can be written as the condition that

$$\rho(X) = \zeta d\pi(X) \quad (X \in \mathfrak{g}_1)$$

be Hermitian symmetric. Thus at the formal level we define a unitary representation of $(G_0, \mathfrak{g})$ to be a triple $(\pi, \gamma, \mathcal{H})$ where π is an even unitary representation in a super Hilbert space $\mathcal{H}$, γ is a representation of $\mathfrak{g}$ that is compatible with the action of G_0 on $\mathfrak{g}$, with $\gamma = d\pi$ on $\mathfrak{g}_0$ and $\rho(X) = \zeta\gamma(X)$ Hermitian symmetric for all $X \in \mathfrak{g}_1$. In other words, unitary representations are triples $(\pi, \rho, \mathcal{H})$ with the appropriate compatibility and hermitian symmetry conditions described above.

To make this definition precise we have to bring in the domains of the operators $\rho(X)$ for $X \in \mathfrak{g}_1$. It is natural to start with the assumption that they should have a common dense domain. Initially we shall suppose that this domain is the space of *all infinitely differentiable vectors for* π_0, but this will be relaxed later and we shall see that even if one starts with a different domain, *the compatibility and symmetry conditions force the operators* $\rho(X)(X \in \mathfrak{g}_1)$ *to be well defined on the space of all infinitely differentiable vectors for* π_0.

We therefore make the following definition. A unitary representation of a super Lie group $(G_0, \mathfrak{g})$ is a triple $(\pi, \rho, \mathcal{H})$ with the following properties:

(i) π is an even unitary representation of G_0 in the super Hilbert space $\mathcal{H}$.
(ii) ρ is a linear map of $\mathfrak{g}_1$ into the subspace of odd endomorphisms of $C^\infty(\pi)$. Here $C^\infty(\pi)$ is the space of infinitely differentiable vectors for π (it is super linear, i.e., it contains the odd and even components of each of its elements).
(iii) ρ satisfies the requirements below:

(a) $\rho(g_0X) = \pi(g_0)\rho(X)\pi(g_0)^{-1}$ $(X \in \mathfrak{g}_1, g_0 \in G_0)$ (compatibility of ρ with π).

(b) $\rho(X)$ with domain $C^\infty(\pi)$ is symmetric for all $X \in \mathfrak{g}_1$. This means that the adjoint $\rho(X)^*$ is an extension of $\rho(X)$ for $X \in \mathfrak{g}_1$.
(c) $-id\pi([X, Y]) = \rho(X)\rho(Y) + \rho(Y)\rho(X)(X, Y \in \mathfrak{g}_1)$ on $C^\infty(\pi)$.

Theorem 1. *If $(\pi, \rho, \mathcal{H})$ is a unitary representation of $(G_0, \mathfrak{g})$, then for any $X \in \mathfrak{g}_1$ the operator $\rho(X)$ with domain $C^\infty(\pi)$ is essentially self adjoint, and*

$$d\pi : X_0 + X_1 \longmapsto d\pi(X_0) + \zeta^{-1}\rho(X_1)(X_0 \in \mathfrak{g}_0, X_1 \in \mathfrak{g}_1)$$

is a super representation of the super Lie algebra $\mathfrak{g}$ in $C^\infty(\pi)$, compatible with π.

The choice of $C^\infty(\pi)$ as the domain for the operators $\rho(X)$, while natural, might seem arbitrary. For instance one might use the *analytic vectors* instead of the differentiable ones. It turns out that this objection is illusory, i.e., any alternative system leads always to a unitary representation in the above sense. This makes it clear that in spite of appearances our definition of a unitary representation of a super Lie group is a very viable and stable notion. To make these remarks more precise let us consider a system $(\pi, \rho, \mathcal{B}, \mathcal{H})$ with the following properties.

(i) $\mathcal{B}$ is a dense super linear subspace of $\mathcal{H}$ invariant under π, π being an even unitary representation of G_0 in the super Hilbert space $\mathcal{H}$. Moreover $\mathcal{B} \subset D(d\pi(Z))$ for all $Z \in [\mathfrak{g}_1, \mathfrak{g}_1]$, $D(A)$ denoting the domain of the operator A. Here for any $Z \in \mathfrak{g}_0$, we write $-id\pi(Z)$ for the unique self adjoint operator in $\mathcal{H}$ such that $\pi(\exp tZ) = \exp td\pi(Z))$ for all $t \in \mathbf{R}$.
(ii) Each $\rho(X)$ for $X \in \mathfrak{g}_1$ is a linear operator in $\mathcal{H}$ with $D(\rho(X)) \supset \mathcal{B}$ such that $X \longmapsto \rho(X)$ is a linear map of $\mathfrak{g}_1$ into $\mathrm{Hom}(\mathcal{B}, \mathcal{H})$. Moreover,

(a) $\rho(X)$ is symmetric for all $X \in \mathfrak{g}_1$.
(b) $\rho(X)\mathcal{B}_i \subset \mathcal{H}_{i+1}(\mathrm{mod}\ 2)$ for all $X \in \mathfrak{g}_1$.
(c) ρ is compatible with π, i.e.,

$$\pi(g)\rho(X)\pi(g)^{-1} = \rho(gX)$$

on $\mathcal{B}$ for $g \in G_0$, $X \in \mathfrak{g}_1$.
(d) $\rho(X)\mathcal{B} \subset D(\rho(Y))$ for all $X, Y \in \mathfrak{g}_1$ and

$$-id\pi([X, Y]) = \rho(X)\rho(Y) + \rho(Y)\rho(X)$$

for all $X, Y \in \mathfrak{g}_1$ on $\mathcal{B}$.

Theorem 2. *Let $(\pi, \rho, \mathcal{B}, \mathcal{H})$ be as above. Then for any $X \in \mathfrak{g}_1$, $\rho(X)$ is essentially self adjoint on $\mathcal{B}$; and $C^\infty(\pi) \subset D(\overline{\rho(X)})$. Let us write $\overline{\rho}(X)$ for the restriction of $\overline{\rho(X)}$ to $C^\infty(\pi)$. Then $(\pi, \overline{\rho}, \mathcal{H})$ is a unitary representation of $(G_0, \mathfrak{g})$. If $(\pi, \rho', \mathcal{H})$ is a unitary representation of $(G_0, \mathfrak{g})$ such that $\mathcal{B} \subset D(\overline{\rho'(X)})$ and $\overline{\rho'(X)}$ restricts to $\rho(X)$ on $\mathcal{B}$ for all $X \in \mathfrak{g}_1$, then $\rho' = \overline{\rho}$.*

Remark. For operators A, B in $\mathcal{H}$ write $A \prec B$ if $D(A) \subset D(B)$ and B coincides with A on $D(A)$. Let A be a densely defined symmetric operator which is essentially self adjoint. Then $\overline{A}$, the closure of A (which exists because A is symmetric), is the unique self adjoint extension of A; moreover, if B is any symmetric operator such that $A \prec B$, we have $\overline{A} = \overline{B}$ so that $A \prec B \prec \overline{A} = \overline{B}$. Thus, as a consequence of the first statement in Theorem 2 we see that extension by closure of $\rho(X)$ is self adjoint and its domain contains $C^\infty(\pi)$. This is the key fact in the theorems.

In the classical theory of representations of semi simple or reductive Lie groups the K-finite vectors play a decisive role. The space of K-finite vectors, although in general a subspace of the space of analytic vectors, is not invariant under the group but only under the Lie algebra. So it would appear useful to have a variant of Theorem 2 where $\mathcal{B}$ is a subspace of $C^\omega(\pi)$, the space of analytic vectors of π, with $\mathcal{B}$ stable under $\mathfrak{g}$ rather than G_0. We then have the following theorem.

Theorem 3. *(i) If $(\pi, \rho, \mathcal{H})$ is a unitary representation of $(G_0, \mathfrak{g})$, then the $\rho(X)$ for $X \in \mathfrak{g}_1$ map $C^\omega(\pi)$ into itself, and $d\pi$ as in Theorem 1 is a representation of $\mathfrak{g}$ in $C^\omega(\pi)$. (ii) Let G_0 be connected, let $\mathcal{B}$ be a dense super linear subspace of $C^\omega(\pi)$, and let $d\pi$ be a representation of $\mathfrak{g}$ in $\mathcal{B}$ such that $d\pi(Z) \prec d\pi_0(Z)$ for all $Z \in \mathfrak{g}_0$ and $\rho(X) = \zeta\pi(X)$ symmetric for $X \in \mathfrak{g}_1$. Then for any $X \in \mathfrak{g}_1$, $\rho(X)$ is essentially self adjoint on $\mathcal{B}$ and $C^\infty(\pi) \subset D(\overline{\rho(X)})$. Let us write $\overline{\rho}(X)$ for the restriction of $\overline{\rho(X)}$ to $C^\infty(\pi)$. Then $(\pi, \overline{\rho}, \mathcal{H})$ is a unitary representation of $(G_0, \mathfrak{g})$. Finally, if $(\pi, \rho', \mathcal{H})$ is a unitary representation of $(G_0, \mathfrak{g})$ such that $\mathcal{B} \subset D(\overline{\rho'(X)})$ and $\rho'(X)$ restricts to $\rho(X)$ on $\mathcal{B}$ for all $X \in \mathfrak{g}_1$, then $\rho' = \overline{\rho}$.*

With these theorems we have realized our objective, namely to introduce a viable *category of unitary representations of a super Lie group*. A morphism from $\Pi = (\pi, \rho, \mathcal{H})$ to $\Pi' = (\pi', \rho', \mathcal{H}')$ is a bounded linear map $A(\mathcal{H} \longrightarrow \mathcal{H}')$ that intertwines π, π' and ρ, ρ'. Note that as soon as A intertwines π and π', it maps $C^\infty(\pi)$ into $C^\infty(\pi')$, and so the statement that A intertwine ρ, ρ' makes sense. If A is an isomorphism and unitary we speak of unitary

equivalence. Π' is a subrepresentation of Π if $\mathcal{H}'$ is a closed graded subspace of $\mathcal{H}$ invariant under π, $\mathcal{H}' \cap C^\infty(\pi)$ is invariant under ρ, and π' (resp. ρ') is the restriction of π (resp. ρ) to $\mathcal{H}'$ (resp. $\mathcal{H} \cap C^\infty(\pi)$). If Π' is a proper nonzero subrepresentation of Π, $\mathcal{H}'' = \mathcal{H}'^{\perp}$, and π'' is the restriction of π to $\mathcal{H}''$, then $C^\infty(\pi)$ is the direct sum od $C^\infty(\pi')$ and $C^\infty(\pi'')$, and it follows from the essential self adjointness of $\rho(X)(X \in \mathfrak{g}_1)$ that $C^\infty(\pi'')$ is stable under all the $\rho(X)$; if ρ'' is the restriction of ρ to $C^\infty(\pi'')$, then $\Pi'' := (\pi'', \rho'', \mathcal{H}'')$ is a unitary representation of $(G_0, \mathfrak{g})$, and $\Pi = \Pi' \oplus \Pi''$. Π is irreducible if the only subrepresentation of Π that is non zero is Π itself. We have the following theorem.

Theorem 4. *(i)* Π *is irreducible if and only if* $\mathrm{Hom}(\Pi, \Pi) = \boldsymbol{C}$. *(ii) Let* Π *be a unitary representation of* $(G_0, \mathfrak{g})$ *and let* P^X *be the spectral measure of* $\overline{\rho(X)}$ *for* $X \in \mathfrak{g}_1$. *If* $\mathcal{H}'$ *is a closed super linear subspace of* $\mathcal{H}'$, *then the following statements are equivalent: (a)* $\mathcal{H}'$ *is stable under* π *and* $\mathcal{H}' \cap C^\infty(\pi)$ *is stable under all* $\rho(X)(X \in \mathfrak{g}_1)$ *(b)* $\mathcal{H}'$ *is stable under* π *and all the projections* P_F^X, F *Borel and* $\subset \boldsymbol{R}$. *In particular,* Π *is irreducible if and only if there is no nonzero proper closed super linear subspace invariant under* π *and all* $P_F^X (X \in \mathfrak{g}_1)$.

Some comments on the proofs

The most striking aspects of the above theorems are the essential self adjointness of the $\rho(X)(X \in \mathfrak{g}_1)$ and the fact that $C^\infty(\pi) \subset D(\overline{\rho(X)})$. It is clearly because of these facts that no matter what subspace $\mathcal{B}$ we start from, all the operators $\rho(X)$ extend naturally to $C^\infty(\pi_0)$ and we obtain a unitary representation in the precise sense we have defined. I shall try to explain the idea behind proving these facts.

First let us introduce some definitions. If A is a closable operator and $\mathcal{D} \subset D(A)$ a dense linear subspace, we say that $\mathcal{D}$ is a *core* for A if A is the closure of its restriction to $\mathcal{D}$. Let A be symmetric. We say that a vector $\psi \in D(A)$ is *analytic for* A if $\psi \in D(A^n)$ for all n and the series $\sum_n t^n ||A^n \psi||/n!$ converges for some $t > 0$. It is a well known result that if $\mathcal{A} \subset D(A)$ contains a dense set of vectors analytic for A then A is essentially self adjoint and $\mathcal{A}$ is a core for A.

In SUSY quantum mechanics the key assumption is that the Hamiltonian $H = L^2$, L being an odd symmetric operator. It is usually suggested that this assumption implies the positivity of the energy operator H. Actually this

condition has deep consequences, not just to the positivity of the energy but to our entire theory. Let us consider the context of Theorem 2. Then, for $X \in \mathfrak{g}_1$, we have on $\mathcal{B}$,

$$2\rho(X)^2 = \rho([X, X]) = -id\pi([X, X]).$$

Now

$$U_t = \pi(\exp t[X, X]) = e^{itK}$$

where K is self adjoint by Stone's theorem. Under our hypotheses

$$K = 2\rho(X)^2$$

on $\mathcal{B}$. The key lemma in the proofs of the theorems stated above is the following.

Lemma. *Let H be a self adjoint operator and $U_t = e^{itH}$. Let $\mathcal{A} \subset D(H)$ be a dense linear subspace. Assume that $\mathcal{A}$ satisfies one of the two conditions: (a) $\mathcal{A}$ is invariant under the one-parameter group (U_t) (b) $\mathcal{A}$ contains a dense set of vectors analytic for H. We then have the following.*

(i) $\mathcal{A}$ is a core for H.

(ii) Let L be a symmetric operator with $\mathcal{A} \subset D(L)$ such that $L\mathcal{A} \subset D(L)$ (so that $\mathcal{A} \subset D(L^2)$) and L^2 coincides with H on $\mathcal{A}$. Then L is essentially self adjoint and $\mathcal{A}$ is a core for it. Moreover, $H = \overline{L}^2$. In particular, $H \geq 0$, $D(H) \subset D(\overline{L})$.

Remark. This lemma not only gives the precise conditions under which we can say that a SUSY Hamiltonian is positive, but also shows how the "odd square root" of the Hamiltonian is itself controlled by the Hamiltonian.

Additional remarks

The basic reference for all of this is [1a]. But, already in 1986, Hirai and Fronsdal had explored the idea of unitary representations of general supergroups in [2]. However their treatment is only formal; there is no discussion of domain considerations for the odd operators, which is a crucial part of the treatment in [1a]. Nevertheless the idea of working with pairs $(G_0, \mathfrak{g})$ can be found in [2]. Part of the reason for the absence of a more precise treatment in [2] must have been the fact that there was no viable definition of super Lie groups available at that time. On the other hand, the physicists' treatment

goes back much earlier to the early 1970's, almost immediately after Salam and Starthdee got the idea of super manifolds (see references [4a, 4b] of the previous essay), but was entirely infinitesimal and confined only to the super Poincaré algebras. I shall make remarks on their results later.

Positive quadratic structures on $\mathfrak{g}_1$

The self adjointness of the $\rho(X)$ implies that $\rho(X)^2$ is a nonnegative self adjoint operators for all $X \in \mathfrak{g}_1$. Thus we have a *positive quadratic structure* on $\mathfrak{g}_1$, and hence a *positive Clifford structure*. Thus one of the basic principles of the theory of unitary representations of super Lie groups already emerges at this level, namely, it is a combination of unitary representation theory of ordinary Lie groups with the representation theory of positive Clifford algebras.

5.2 Super imprimitivity theorem for even homogeneous spaces

For ordinary Lie groups the central results of representations of *semi direct product Lie groups* follow as an application of the *imprimitivity theorem* which establishes a functorial equivalence between the category of unitary representations of a closed subgroup H_0 of the Lie group G_0 and systems of imprimitivity for G_0 based on $\Omega = G_0/H_0$. Actually the imprimitivity theorem for finite groups goes back to Frobenius. In 1939, Wigner [3] developed an extension of the Frobenius theory for the Poincaré group, and then, in the late 1940s and early 1950s, Mackey proved a far-reaching generalization of the Frobenius theory, valid in the enormously much wider category of locally compact second countable groups; see [4]. Mackey's theory was applicable to the semi direct products and contained the Wigner theory as a special case. We are interested only in the case of real Lie groups. In the next essay we shall discuss these ideas when spacetime is nonarchimedean so that the groups involved are p-adic Lie groups.

Let G_0 be any locally compact second countable group and Ω a locally compact second countable space on which G_0 acts. A *system of imprimitivity* (SI) for G_0 *based on* Ω is nothing more than a representation of G_0 together with data that encode the G_0-action on Ω. More precisely it is a pair (π, P) that consists of a unitary representation π of G_0 in a Hilbert space $\mathcal{H}$ and a

projection valued measure P on Ω in $\mathcal{H}$ such that

$$\text{(SI)} \qquad \pi(g)P_E\pi(g)^{-1} = P_{gE} \quad (g \in G_0, E \text{ Borel } \subset \Omega).$$

Apart from its intrinsic interest it is also a notion that is essential in formulating the concept of covariance of quantum systems, as was first emphasized by Mackey. Suppose in fact that Ω is the *configuration space* of a quantum system, and we want to formulate the idea that our description is to be symmetric with respect to G_0. The functions on Ω are the "coordinates" that determine the "configurations" of the system considered and so we should require that we have "position operators" that correspond to them. This means that for each (Borel) subset $E \subset \Omega$ we have a projection operator P_E in the quantum Hilbert space $\mathcal{H}$ the defines the experiment that decides whether the configuration of the system is located in E. The assignment

$$E \longmapsto P_E$$

should be a projection valued measure. Then covariance with respect to G_0 means that there is a unitary representation U of G_0 in $\mathcal{H}$ such that

$$U_g P_E U_g^{-1} = P_{gE}$$

For instance, if $\Omega = \mathbf{R}^n$ and G_0 is the group of translations this leads to the Heisenberg-Weyl systems. From the point of view of quantum mechanics it is of great importance to know how to describe *all* systems of imprimitivity for a given (Ω, G_0).

We can also make a smooth version of this notion if we replace P by its smooth version, namely, a $*$-homomorphism

$$f \longmapsto A(f)$$

of the algebra $C_c^\infty(\Omega)$ of smooth compactly supported functions on Ω into the $*$-algebra of bounded operators in $\mathcal{H}$, such that

$$U_g A(f) U_g^{-1} = A(g[f]) \qquad (g \in G, f \in C_c^\infty(\Omega)),$$

where $g[f]$ is the transform of the function f by the group element g.

Actually, one can make a vast generalization of this as follows. We think of Ω *not* as a configuration space but as "spacetime." For each open set $E \subset \Omega$ we suppose that have a $*$-homomorphism

$$f \longmapsto A(f) \qquad f \in C_c^\infty(E)$$

into the algebra of bounded operators in $\mathcal{H}$, which encode the measurements we carry out in E (which are in the *quantum algebra* associated to E). Then the covariance is formulated as the relation

$$U_g A(f) U_g^{-1} = A(g[f]),$$

where $g[f]$ is the transform of the function f by the group element g. The $A(f)$ will not in general commute with each other but will do so, if the supports of f and g are *space-like separated* in the relativistic sense, namely, when light signals from one of them cannot reach the other. Then such a system will be the starting point of a *quantum field theory*, as was first discovered by Wightman. The fact that the operators $A(f)$ commute only in a very restricted fashion makes the determination of such systems extremely hard.

Let us come back to systems of imprimitivity. Even here we have to do some preliminary reduction because G_0 may not act *transitively* on Ω. Since the conditions combine nicely when we form the direct sum of two systems associated to two actions of G_0, it follows heuristically that any system is a direct sum (or integral) of systems that correspond to the *orbits* where the action is *transitive*. This is in fact an over simplification as it ignores the existence of *ergodic actions*. For instance, the classic Lebesgue action of $\mathbf{Z}$ on the circle T defined by an irrational rotation is not transitive but ergodic. In any case, one still has to understand systems of imprimitivity based on transitive actions before going on to treat ergodic systems.

For a *transitive* action the story is very simple: if H_0 is the stabilizer of some point ω_0 of Ω, the category of transitive systems on Ω is functorially equivalent to the category of unitary representations of H_0. This is Mackey's great theorem, which is a culmination of a line of thought that started with Frobenius, and evolved through the ideas of von Neumann, Wigner, Gel'fand, and possibly others.

If σ is a unitary representation of H_0 in a Hilbert space $\mathcal{K}^\sigma$, then one can associate canonically to σ a SI (π^σ, P^σ) for G_0 based on Ω: π^σ is the representation of G_0 *induced by* σ and P^σ is the natural projection valued measure on Ω in the space of π^σ. Here we must recall that the Hilbert space $\mathcal{H}^\sigma$ of π^σ is the space of (equivalence classes of) measurable functions f from G_0 to $\mathcal{K}^\sigma$ such that

(a) $f(x\xi) = \sigma(\xi)^{-1} f(x)$ for each $\xi \in H_0$ for almost all x and
(b) $\int |f(x)|^2_{\mathcal{K}^\sigma} d\bar{x} < \infty$ (in view of (a) the $\mathcal{K}^\sigma$-norm of $f(x)$ is really defined on Ω, and $d\bar{x}$ is the invariant measure* on Ω.).

The representation π^σ acts by left translation. For any Borel $E \subset \Omega$ the projection P^σ_E is the map $\psi \longmapsto \chi_E \psi$ ($\psi \in \mathcal{H}^\sigma$) where χ_E is the characteristic function of E. For the smooth version, where P^σ is replaced by A^σ, the operator $A^\sigma(f)$ is multiplication by f: $\psi \longmapsto f\psi$. The imprimitivity theorem is the statement that

$$\sigma \longmapsto (\pi^\sigma, P^\sigma) \qquad \text{or} \qquad \sigma \longmapsto (\pi^\sigma, A^\sigma)$$

is a functorial equivalence from the category of unitary representations of H_0 to the category of SIs based on Ω. The representation π^σ is said to be *induced* by σ.

The above definition is not very intuitive and I shall now describe induced representations in a more heuristic manner. We think of $\mathcal{K}^\sigma$ as "attached" to the point ω_0 which defines the stabilizer H_0. If $\omega = g[\omega_0]$, we attach a Hilbert space $\mathcal{K}^\sigma_\omega$ to ω and a unitary isomorphism $L_g : \mathcal{K}^\sigma \longrightarrow \mathcal{K}^\sigma_\omega$. In this manner we have a "vector bundle" on Ω on which G_0 acts. The induced representation is then the natural action of G_0 on the space of *sections* of this bundle. One can make this rigorous as follows. We take $G_0 \times \mathcal{K}^\sigma$ on which G_0 acts from the left, and introduce the equivalence relation

$$(g, \psi) \sim (g\xi, \sigma(\xi)^{-1}\psi).$$

It is easy to see that the set

$$\mathcal{V}^\sigma := G_0 \times \mathcal{K}^\sigma / \sim$$

of equivalence classes can be interpreted as a vector bundle over Ω whose fibers are isomorphic to $\mathcal{K}^\sigma$. The action of G_0 on $G_0 \times \mathcal{K}^\sigma$ comes down to an action on $\mathcal{V}^\sigma$ and hence leads to an action on the space of sections of $\mathcal{V}^\sigma$. The sections acquire a natural L^2-norm and the space $\mathcal{H}^\sigma$ is just the subspace of square integrable sections. If G_0 is a Lie group and $\mathcal{K}^\sigma$ is finite dimensional (not necessarily unitary), the assignment

$$\mathcal{K}^\sigma \longmapsto \mathcal{V}^\sigma$$

* The theory can be developed even if there is no invariant measure, but for most applications invariant measures are available.

is just the classical functor that goes from the category of H_0-modules to the category of vector bundles with a G_0-action. Important geometric examples are the various tensor and spinor bundles on Ω. It must be emphasized that the functor

$$\sigma \longmapsto \pi^\sigma$$

is *not* an equivalence of categories in general, but the functor

$$\sigma \longmapsto (\pi^\sigma, P^\sigma)$$

is, according to the Mackey theorem.

It is well known how this theorem leads in a straightforward manner to the famous Mackey-Wigner classification of irreducible unitary representations (UIRs) of semidirect products $G_0 = T_0 \times' L_0$ where T_0 is an abelian group with an action $h, a \longmapsto h[a]$ of L_0 on it. A UR W of G_0 gives rise to pair (U, V) of URs of T_0 and L_0 by restriction which are coupled by the relation

$$V(h)U(a)V(h)^{-1} = U(h[a]) \qquad (a \in T_0, h \in L_0).$$

We diagonalize U by expressing it as an integral with respect to a projection valued measure P on the dual group $\widehat{T_0}$ of T_0, and the above relation translates into

$$V(h)P_E V(h)^{-1} = P_{h[E]} \qquad (h \in L_0, E \text{ Borel } \subset \widehat{T_0}).$$

The support of P is the *spectrum* of W. If W is irreducible, P is *ergodic* in the sense that for any H-*invariant* Borel set E above, $P_E = 0$ or I. Indeed, as $h[E] = E$, P_E will commute with V by the above relation, while it commutes with U obviously, and so will commute with W. Suppose we assume that P is concentrated in a *single orbit* for L_0. Then by restricting to that orbit we obtain a transitive system of imprimitivity. If χ is a point in the supporting orbit and L_χ is the stabilizer of χ in L_0, we thus see that the (equivalence classes of) UIRs of L_χ classify up to equivalence the UIRs of G_0 whose spectrum is the given orbit. If the representation of L_χ is irreducible, so will the representation for G_0, and vice versa.

For the general case it might happen that there are ergodic systems whose spectrum is not concentrated in a single orbit. In this case there will be many more UIRs of G_0 than are accounted for by UIRs of the various L_χ. This is however a situation not encountered in many natural applications. Suppose now we assume that the semidirect product is *regular*; this means that the action of L_0 on $\widehat{T_0}$ is nice in the Borel sense, i.e., there is a Borel cross section

for the orbits. In this case all ergodic systems have spectra concentrated in orbits. Every UIR of G_0 is then obtained by this method, by varying the orbit and the UIR of the L_χ. The groups L_χ are the *little groups* of the physics literature that go back to Wigner when he classified the UIRs of the Poincaré group in his famous paper in 1939 [3]. The general theory described above is due to Mackey [4].

Effros has proved the beautiful theorem that under very general conditions semidirect products are regular, for instance if the L_0-orbits in $\widehat{T}_0$ are all locally closed, namely, are open in their closures [5]. This is the case for instance when the actions are *algebraic.*

I shall now formulate a generalization of the imprimitivity theorem to the super setting. This will lead to a classification of irreducible unitary representations of super semi direct products. To this end we work with a super Lie group $G = (G_0, \mathfrak{g})$ and a closed super subgroup $H = (H_0, \mathfrak{h})$. We shall assume that the subgroup is *special*, namely that $\mathfrak{h}_1 = \mathfrak{g}_1$, or that *the odd part of* $\mathfrak{h}$ *coincides with the odd part of* $\mathfrak{g}$. This means that the *associated homogeneous space is classical, namely* $\Omega := G/H = G_0/H_0$. It is an interesting problem to remove this restriction. However this condition is adequate enough to provide a solution to our main problem, that of classifying the irreducible unitary representations of regular super semidirect products. *All* the results in the physical literature are contained in this treatment since all super Poincaré groups and their variants encountered in physics are regular super semidirect products.

The super imprimitivity theorem

Let $H = (H_0, \mathfrak{h})$ be a special subgroup of $G = (G_0, \mathfrak{g})$. A *super system of imprimitivity* (SSI) for $(G_0, \mathfrak{g})$ based on $\Omega = G_0/H_0$ is a system $(\pi, \rho^\pi, \mathcal{H}, P)$ with the following properties:

(1) The triple $(\pi, \rho^\pi, \mathcal{H})$ is a unitary representation of the super Lie group $(G_0, \mathfrak{g})$

(2) The triple $(\pi, \mathcal{H}, P)$ is a classical system of imprimitivity for G_0 in $\mathcal{H}$, based on Ω, P being even

(3) The projection valued measure P commutes with ρ^π; this means that the spectral projections of the $\rho^\pi(X)(X \in \mathfrak{g}_1)$ commute with the projections $P_E(E \text{ Borel } \subset \Omega)$.

The last condition, which is the new ingredient in the super setting, may be explained as follows. The projections of P give rise to $*$-homomorphism $f \longmapsto A(f)$ of the commutative algebra of bounded Borel functions on Ω into the algebra of bounded even operators on $\mathcal{H}$. The additional requirement of compatibility between the odd part of the super representation and the projections of P may be formulated infinitesimally by saying that

$$[\rho^{\pi}(X), A(f)] = A(X(f)) \qquad (X \in \mathfrak{g}_1, f \in C_c^{\infty}(\Omega));$$

but, as $\Omega = G/H = G_0/H_0$ is *purely even*, $Xf = 0$ when X is odd and f is even. This is encoded by (3).

We shall now associate to any unitary representation $(\sigma, \rho^{\sigma}, \mathcal{K}^{\sigma} = \mathcal{K})$ of the super Lie group $(H_0, \mathfrak{h})$, a SSI of $(G_0, \mathfrak{g})$. To this end we must define π and ρ^{π}. We shall use the flexibility provided by Theorem 2 above and define the $\rho^{\pi}(X)$ on a suitable space $\mathcal{B}$. We define π by

$$\pi = \mathrm{Ind}_{H_0}^{G_0}\sigma$$

in the Hilbert space $\mathcal{H}$ introduced earlier. The super Hilbert space structure of $\mathcal{K}$ gives rise to a super Hilbert space structure for $\mathcal{H}$ and π becomes even. Let $\mathcal{B} = C_c^{\infty}(\pi)$, the space of C^{∞} vectors for π in $\mathcal{H}$ which, as functions from G_0 to $\mathcal{K}$, are smooth and have compact support mod H_0. It can be shown (as a consequence of the Dixmier-Malliavin theorem) that $\mathcal{B}$ is precisely the space of all smooth functions f from G_0 to $\mathcal{K}$ such that $f(x\xi) = \sigma(\xi)^{-1} f(x) (x \in G_0, \xi \in H_0)$ and have compact support mod H_0. It is then easy to see that the values of any $f \in \mathcal{B}$ lie in $C^{\infty}(\sigma)$ so that all the operators $\rho^{\sigma}(X) (X \in \mathfrak{h}_1 = \mathfrak{g}_1)$ act on all the values of any $f \in \mathcal{B}$. We define

$$(\rho^{\pi}(X)f)(x) = \rho^{\sigma}(x^{-1}X)f(x) \quad (f \in \mathcal{B}, x \in G_0).$$

Of course the projection valued measure is the standard one associated to π. We then have the following theorem [1a].

Theorem 5. *The quadruple $(\pi, \rho^{\pi}, \mathcal{B}, P)$ is a SSI for the super Lie group $(G_0, \mathfrak{g})$ based on $\Omega = G_0/H_0$. The assignment*

$$(\sigma, \rho^{\sigma}, \mathcal{K}) \longmapsto (\pi, \rho^{\pi}, P)$$

is a functorial equivalence of categories from the category of unitary representations of $(H_0, \mathfrak{h}_1)$ to the category of SSIs for $(G_0, \mathfrak{g})$ based on $\Omega = G_0/H_0$. In

particular we have an equivalence preserving bijection from UIRs of $(H_0, \mathfrak{h})$ *with irreducible SSIs for* $(G_0, \mathfrak{g})$ *based on* G_0/H_0.

Remark. There are quite a few technical details to be verified in the proof of the Theorem. I refer the reader to [1a].

5.3 Super semidirect products and their unitary irreducible representations

There are several places in the above treatment where we have made new definitions based on heuristic considerations. It is now essential now to discuss applications and make sure that our definitions are justified at least *ex posto facto*. We shall do this by discussing UIRs of super semi direct products.

We start with a *classical* semi direct product $G_0 = T_0 \times' L_0$ where T_0 is a real finite dimensional vector space (the spacetime translation group) and L_0 a closed unimodular subgroup of $GL(T_0)$ acting naturally on T_0. We shall assume that the semi direct product is *regular*, namely, that for the action of L_0 on the dual T_0^* there is a Borel cross section; this is equivalent to saying that all the L_0-orbits in T_0^* are locally closed (Effros). Let $\mathfrak{t}_0 = \mathrm{Lie}(T_0)$, $\mathfrak{l}_0 = \mathrm{Lie}(L_0)$, $\mathfrak{g}_0 = \mathrm{Lie}(G_0)$.

I have remarked that in the construction of super Lie algebras $\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1$ the difficult condition to satisfy is the cubic nonlinearity

$$[X, [X, X]] = 0 \qquad (X \in \mathfrak{g}_1).$$

One way to take care of this is to ensure that the odd bracket goes into a subspace $\mathfrak{g}_2 \subset \mathfrak{g}_0$ which *acts trivially* on $\mathfrak{g}_1$:

$$[\mathfrak{g}_2, \mathfrak{g}_1] = 0$$

The triple bracket condition above is then automatic. In what follows we must remember this.

By a *super translation group* we mean a super Lie group $(T_0, \mathfrak{t})$ where $\mathfrak{t}_0$ is abelian and acts trivially on $\mathfrak{t}_1$. Note that as $[\mathfrak{t}_1, \mathfrak{t}_1] \subset \mathfrak{t}_0$ and $\mathfrak{t}_0$ acts trivially on $\mathfrak{t}_1$, the condition for a super Lie algebra is automatic. Suppose now that $\mathfrak{t}_1$ is an L_0-module and that the super commutator map $a, b \longmapsto [a, b]$ is L_0-equivariant from $\mathfrak{t}_1 \times \mathfrak{t}_1$ into $\mathfrak{t}_0$. Then $\mathfrak{g} := \mathfrak{l}_0 \oplus \mathfrak{t}$ is a super Lie algebra with $\mathfrak{g}_0 = \mathfrak{l}_0 \oplus \mathfrak{t}_0 = \mathrm{Lie}(G_0)$, $\mathfrak{g}_1 = \mathfrak{t}_1$, for the same reasons as for $\mathfrak{t}$, namely, that

the odd bracket goes into $\mathfrak{t}_0$ which acts trivially on $\mathfrak{t}_1$. The super Lie group $S := (G_0, \mathfrak{g})$ is the *super semi direct product* of L_0 and the super translation group $(T_0, \mathfrak{t})$. In this manner we have obtained a whole class of super Lie groups (which are algebraic if L_0 and its actions on $\mathfrak{t}_0$ and $\mathfrak{t}_1$ are algebraic); the data needed for their construction are L_0, and L_0-modules $\mathfrak{t}_0$, $\mathfrak{t}_1$, together with a symmetric L_0-equivariant bilinear map of $\mathfrak{t}_1 \times \mathfrak{t}_1 \longrightarrow \mathfrak{t}_0$.

For any closed subgroup $S_0 \subset L_0$ we have $H_0 = T_0 S_0$ is a closed subgroup of G_0 and $(H_0, \mathfrak{h})$ is a *special* super Lie subgroup of $(G_0, \mathfrak{g})$ where $\mathfrak{h} = \mathfrak{h}_0 \oplus \mathfrak{t}_1$, $\mathfrak{h}_0$ being $\mathrm{Lie}(H_0)$; if $\mathfrak{s}_0 = \mathrm{Lie}(S_0)$, then $\mathfrak{h}_0 = \mathfrak{s}_0 \oplus \mathfrak{t}_0$. The super Poincaré groups that occur in the physical literature are special cases of the above construction where L_0 is an orthogonal group (rather its two-fold cover, the *spin group*) with respect to a nondegenerate quadratic form on $\mathfrak{t}_0$, and $\mathfrak{t}_1$ is chosen as a *spin module* for L_0. But in this section we shall discuss the general case. We are interested in describing *all* the unitary irreducible representations of $(G_0, \mathfrak{g})$.

The aim is to show that the irreducible unitary representations of the super semi direct product may be classified by a straightforward extension of the Frobenius-Mackey-Wigner theory, namely by the *method of little super groups*. For any $\lambda \in T_0^*$ let L_0^λ be the stabilizer of λ in L_0 and let

$$\mathfrak{g}^\lambda = \mathfrak{t}_0 \oplus \mathfrak{l}_0^\lambda \oplus \mathfrak{g}_1, \quad \mathfrak{l}_0^\lambda = \mathrm{Lie}(L_0^\lambda).$$

We shall refer to the super Lie group $S^\lambda := (T_0 L_0^\lambda, \mathfrak{g}^\lambda)$ as the *little super group at* λ. It is a special sub super Lie group of $S = (G_0, \mathfrak{g})$. Given a unitary representation (σ, ρ^σ) of S^λ, we shall say that it is λ-*admissible* if $\sigma(t) = e^{i\lambda(t)} I$ for all $t \in T_0$; λ itself is called *admissible* if there is an irreducible unitary representation of S^λ which is λ-admissible. Let

$$T_0^+ = \left\{ \lambda \in T_0^* \,\middle|\, \lambda \text{ admissible} \right\}.$$

It is easy to see that T_0^+ is L_0-invariant.

Given a unitary representation (π, ρ^π) of the super Lie group S we obtain a spectral measure on P on T^* by Fourier transformation of the restriction of π to T_0:

$$\pi(t) = \int_{T^*} e^{i\lambda(t)} dP(\lambda) \quad (t \in T).$$

If E is any L_0-invariant Borel set, the spectrum of the unitary representation is said to be *in* E if $P_E = I$.

Theorem 6. *The spectrum of every irreducible unitary representation of the super Lie group $S = (G_0, \mathfrak{g})$ is in some orbit in T_0^+. For each orbit in T_0^+ and choice of λ in that orbit, the assignment that takes a λ-admissible unitary representation $\gamma := (\sigma, \rho^\sigma)$ of S^λ into the unitary representation U^γ of $(G_0, \mathfrak{g})$ induced by it, is a functor which is an equivalence of categories between the category of the λ-admissible unitary representations of S^λ and the category of unitary representations of $(G_0, \mathfrak{g})$ with their spectra in that orbit. Varying λ in that orbit changes the functor into an equivalent one. In particular this functor gives a bijection between the respective sets of equivalence classes of irreducible unitary representations.*

Proof (Sketch). Since T_0 acts trivially on $\mathfrak{g}_1$ it follows that for any unitary representation (π, ρ^π) of S the operators $\pi(t)(t \in T_0)$ commute with $\rho^\pi(X)(X \in \mathfrak{g}_1)$ and so the spectral measure P and ρ^π commute with each other. If (π, ρ^π) is irreducible and E is a L_0-invariant Borel set, P_E commutes with π also, hence with (π, ρ^π), and so $P_E = 0$ or I. Thus, by regularity of the semi direct product, $P_O = I$ for some orbit O. If $\lambda \in O$ we can view P as a projection valued measure on L_0/L_0^λ. Clearly (π, ρ^π, P) is a SSI for S based on L_0/L_0^λ. By the super imprimitivity theorem this is the SSI associated to a unitary representation (σ, ρ^σ) of S^λ. From the classical theory one knows that $\sigma(t) = e^{i\lambda(t)}I$ for all $t \in T_0$. If (π, ρ^π) is irreducible, (σ, ρ^σ) is irreducible, and so λ is admissible, i.e., $\lambda \in T^+$.

Remark 1. The reader should notice the sharp difference between the classical and super symmetric situations. In the classical theory, *all* orbits in T_0^* are allowed, while in the SUSY case only those in T_0^+ are. Actually, when everything is even, $T_0^+ = T_0^*$; in fact, given any λ, the map $tx \longmapsto e^{i\lambda(t)}(t \in T_0, x \in L_0^\lambda)$ is well defined and gives an one dimensional irreducible unitary representation of $T_0 L_0^\lambda$, thus showing that $\lambda \in T_0^+$. But when super symmetry is present, the restriction to T_0^+ is a genuine *selection rule*. We shall see below that T_0^+ may be interpreted as the condition of the *positivity of energy*. This is very different from the classical theory where representations of negative energy (tachyons) are theoretically possible and have to be excluded *ad hoc*.

Remark 2. One other point should be noticed. If we restrict the UIR of $(G_0, \mathfrak{g})$ to G_0 we obtain a UR of G_0 that is not in general irreducible. In principle the irreducibles in that restriction may belong to different orbits. But the theorem shows that all the irreducibles in that restriction *belong to the same orbit*. The reason for this is that the odd operators commute with the translations. This

means that all particles in a super multiplet, the so-called *partners*, *have the same mass.* Of course supersymmetry breaking will eventually endow the partner particles with differing masses.

Suppose $\lambda \in T_0^*$ and (σ, ρ^σ) is a unitary representation of S^λ such that $\sigma(t) = e^{i\lambda(t)}I$ for all $t \in T_0$. Then

$$-id\sigma(Z) = \lambda(Z)I \quad (Z \in \mathfrak{t}_0).$$

On the other hand, if $X_1, X_2 \in \mathfrak{g}_1$, then $[X_1, X_2] \in \mathfrak{t}_0$. Thus

$$[\rho^\sigma(X_1), \rho^\sigma(X_2)] = 2\Phi_\lambda(X_1, X_2)I \quad (X_1, X_2 \in \mathfrak{g}_1)$$

where Φ_λ is the *symmetric* bilinear form on $\mathfrak{g}_1 \times \mathfrak{g}_1$ defined by

$$\Phi_\lambda(X_1, X_2) = (1/2)\lambda([X_1, X_2]).$$

If we now take $X_1 = X_2 = X$ we see that

$$\rho^\sigma(X)^2 = Q_\lambda(X)I, \quad Q_\lambda(X) = \Phi_\lambda(X, X) \quad (X \in \mathfrak{g}_1)$$

on $C^\infty(\sigma)$. Since $\rho^\sigma(X)$ is essentially self adjoint on $C^\infty(\sigma)$ it follows that

$$Q_\lambda(X) \geq 0 \quad (X \in \mathfrak{g}_1).$$

This is thus a necessary condition that there should exist a unitary representation (σ, ρ^σ) with $\sigma(t) = e^{i\lambda(t)}I$ for all $t \in T_0$.

We also note that Q_λ is invariant under L_0^λ because for $X_1, X_2 \in \mathfrak{g}_1$, $h \in L_1$,

$$[\rho^\sigma(hX_1), \rho^\sigma(hX_2)] = \sigma(h)[\rho^\sigma(X_1), \rho^\sigma(X_2)]\sigma(h)^{-1} = \Phi_\lambda(X_1, X_2).$$

Theorem 7. *For a $\lambda \in T_0^*$ the following are equivalent.*

(a) There is an irreducible unitary representation (σ, ρ^σ) of S^λ with $\sigma(t) = e^{i\lambda(t)}I$ for all $t \in T_0$, that is, that λ is admissible.

(b) There is a unitary representation (σ, ρ^σ) of S^λ with $\sigma(t) = e^{i\lambda(t)}I$ for all $t \in T_0$.

(c) $Q_\lambda(X) \geq 0$ for all $X \in \mathfrak{g}_1$.

Remark. When the super semi direct product is specialized to the super Poincaré group, the condition $Q_\lambda \geq 0$ is the same as the *positivity of the energy*. This is the reason why we refer to this condition as the positive energy condition in the general case.

Clifford structure when $Q_\lambda \geq 0$

To complete the proof of Theorem 2 we must show that if $Q_\lambda \geq 0$ then there is an irreducible unitary representation (σ, ρ^σ) of S^λ such that $\sigma(t) = e^{i\lambda(t)}I$ for all $t \in T_0$. Actually it is necessary to construct all such so that we can get a description of all irreducible unitary representations of the super semi direct product $S = (G_0, \mathfrak{g})$. This depends on exploiting the Clifford structure that arises from the quadratic form Q_λ. The form Q_λ is invariant under L_0^λ. Moreover the condition $\rho^\sigma(X)^2 = Q_\lambda(X)I$ shows that the operators $\rho^\sigma(X)$ are *bounded self adjoint* so that the situation is much nicer than for the big group. We define $\mathcal{C}_\lambda$ as the quotient of the tensor algebra generated by $\mathfrak{g}_1$ with respect to the relations $X^2 = Q_\lambda(X)(X \in \mathfrak{g}_1)$. If Q_λ is non degenerate this would be the Clifford algebra associated to the quadratic form Q_λ; by abuse of language we shall refer to it as the Clifford algebra associated to Q_λ even if Q_λ is singular.

Clearly ρ^σ may now be viewed as a representation of $\mathcal{C}_\lambda$ in a super Hilbert space $\mathcal{K}$ such that $\rho^\sigma(X)$ is odd, bounded, and self adjoint for all $X \in \mathfrak{g}_1$ (SA representations). Now L_0^λ acts on $\mathfrak{g}_1$ and leaves Q_λ invariant. Hence its action lifts to an action $x, a \longmapsto x[a]$ on $\mathcal{C}_\lambda$. Then the unitary representations of S^λ are pairs (σ, τ) where τ is a SA representation and σ a unitary representation of $T_0 L_0^\lambda$ in the space of τ such that

$$\sigma(t) = e^{i\lambda(t)}(t \in T_0), \quad \tau(x[a]) = \sigma(x)\tau(a)\sigma(x)^{-1} \text{ for } a \in \mathcal{C}_\lambda, x \in L_0^\lambda.$$

Description of the UIRs of S^λ

While there is nothing really difficult in the determination of the pairs (σ, τ), it is technically involved and I do not want to go into all the details. This is because we are in the context of a completely general super semi direct product. One of the difficulties is that L_0^λ need not be connected. Another is that $\mathfrak{g}_1$ mod the radical of Q_λ may have an odd dimension. Neither of these arises for super Poincaré groups. *Let us assume that L_0^λ is connected.*

Since Q_λ may have a radical, say $\mathfrak{r}_\lambda$, we introduce

$$\mathfrak{g}_{1\lambda} = \mathfrak{g}_1/\mathfrak{r}_\lambda.$$

Then Q_λ defines a non degenerate quadratic form on $\mathfrak{g}_{1\lambda}$ which we denote by $Q_\lambda^\sim$. Since L_0^λ preserves Q_λ and is connected, we have a map

$$j : L_0^\lambda \longrightarrow \mathrm{SO}(\mathfrak{g}_{1\lambda}).$$

Then there are two possibilities: (1) The map j lifts to a map (this is the case when L_0^λ is simply connected)

$$\widehat{j} : L_0^\lambda :\longrightarrow \mathrm{Spin}(\mathfrak{g}_{1\lambda}).$$

(2) j does not lift as in (1) but there is a two-fold cover H^λ of L_0^λ such that j lifts to a map

$$\widehat{j} : H^\lambda :\longrightarrow \mathrm{Spin}(\mathfrak{g}_{1\lambda}).$$

In case (2) the covering map

$$H^\lambda \longrightarrow L_0^\lambda$$

has kernel $\{\pm 1\}$. We write ξ for the non trivial element of this kernel.

Lemma 1. *We have the following.*

(a) There exist irreducible SA representations τ of $\mathcal{C}_\lambda$. These are finite dimensional, unique if $\dim(\mathfrak{g}_{1\lambda})$ *is odd, unique up to parity reversal if* $\dim(\mathfrak{g}_{1\lambda})$ *is even. Any SA representation of $\mathcal{C}_\lambda$ is of the form $1 \otimes \tau$.*

(b) Let τ_λ be an irreducible SA representation of $\mathcal{C}_\lambda$. Then, assuming that L_0^λ is connected, there is an even UR κ_λ of H^λ, unique up to multiplication by a character of H^λ, such that

$$\tau_\lambda(x[a]) = \sigma(x)\tau_\lambda(a)\sigma(x)^{-1} \text{ for } a \in \mathcal{C}_\lambda, x \in H^\lambda.$$

The necessary and sufficient condition for κ_λ to be a UR of L_0^λ is that either j lifts to a map $L_0^\lambda \longrightarrow \mathrm{Spin}(\mathfrak{g}_{1\lambda})$ or there exists a character χ of H^λ with $\chi(\xi) = -1$. This is the case if

(C) $\qquad L_0^\lambda = A \times' T, \qquad$ *A is simply connected, T is a torus..*

Moreover κ_λ is unique up to multiplication by a character of L_0^λ.

Proof (Sketch). (a) is a variant of the standard theory of Clifford algebras. Let us now look at (b). The spin group $\mathrm{Spin}(\mathfrak{g}_{1\lambda})$ is imbedded inside the even part of the Clifford algebra of $\mathfrak{g}_{1\lambda}$ and τ_λ may be viewed as a representation of this Clifford algebra. Its restriction, u say, to the spin group is than an even unitary representation of the spin group compatible with the action of $\mathfrak{g}_{1\lambda}$. If

now L_0^λ is simply connected, or more generally, if j lifts to a map $\widehat{j}$ as in case (1), we may take $\kappa_\lambda = u \circ u$. If $\widehat{j}$ exists only on a two-fold cover, $\kappa'_\lambda = u \circ \widehat{j}$ defines the unitary representation we seek. It may happen that $\kappa'_\lambda(\xi) = -1$ (it is ± 1); in this case changing κ'_λ to $\chi\kappa'_\lambda$ we get a unitary representation which descends to L_0^λ, which we may take as κ_λ. Finally, if L_0^λ satisfies (**C**), we can take $H^\lambda = A \times' T^\sim$ where $T^\sim$ is a two-fold cover of T, and $\xi = (1, t)$ where $t \in T^\sim, t \neq 1, t^2 = 1$. There exists a character χ of $T^\sim$ which is -1 at t, and we can extend χ to H^λ by making it trivial on A.

Theorem 8. *Let λ be such that $Q_\lambda \geq 0$ and suppose that condition* (**C**) *on L_0^λ is satisfied. Let $\rho_\lambda = \tau_\lambda\big|_{\mathfrak{g}_1}$, and let $\sigma_\lambda = e^{i\lambda}\kappa_\lambda$ be the unitary representation of $T_0 L_0^\lambda$. We then have the following.*

(*a*) *$(\sigma_\lambda, \rho_\lambda)$ is an irreducible unitary representation of the little super group S^λ which restricts to $e^{i\lambda}$ on T_0.*

(*b*) *The unitary representations (σ, ρ^σ) restricting to $e^{i\lambda}$ on T_0 are precisely those of the form*

$$\sigma = \theta \otimes e^{i\lambda}\kappa_\lambda, \quad \rho^\sigma = 1 \otimes \rho_\lambda$$

where θ is a purely even unitary representtaion of L_0^λ.

(*c*) *The assignment*

$$\theta \longmapsto (\theta \otimes e^{i\lambda}\kappa_\lambda, 1 \otimes \rho_\lambda)$$

is an equivalence of categories from the category of purely even unitary representations of L_0^λ to the category of unitary representations of the little super group S^λ which restrict to $e^{i\lambda}$ on T_0.

Remark. The pair $(e^{i\lambda}\kappa_\lambda, \rho_\lambda)$, which corresponds to the case of trivial θ, is called the *fundamental representation* of S^λ. Once it is determined, the whole structure of the irreducible unitary representations of the original super semi direct product S is made transparent. The determination of κ_λ and the verification of the condition (**C**) are thus the crucial ingredients that remain to be described in the super Poincaré case.

Generalizations

It is a natural question to ask whether we can develop the theory of infinite dimensional representations of more general classes of super Lie groups. There have been many cases treated in the literature, much prior to [1a].

Jakobsen [11] constructed all highest weight unitary modules for classical *super Lie algebras*, while Dobrev et al [12] studied the representations of the conformal supergroups. Recently Salmasian [13] has extended the entire Kirillov theory and obtained a complete theory of unitary representations of nilpotent super Lie groups. It is clear that there is a whole lot to be done in these directions.

5.4 Super Poincaré Lie algebras and Lie groups

It is probably a good idea now to look at the groups that are studied extensively by the physicists. These are the various super Poincaré groups and algebras and variants of them with additional charges, central or otherwise.

The Poincaré group is a semi direct product of the group of spacetime translations and the Lorentz group. For applying supersymmetry to physical problems such as the structure and classification of particles with supersymmetry it is very useful to have supersymmetric extensions of the translation and Poincaré groups. We begin with the translation group. We use the theorem of super Harish-Chandra pairs and think of a super Lie group G as a super Harish-Chandra pair $(G_0, \mathfrak{g})$.

We have already defined the super semidirect products. The super Poincaré groups occurring in physics are very much more special than these. First of all $\mathfrak{t}_0$ is required to have a nondegenerate symmetric quadratic form, namely, a metric, and L_0 must be the corresponding orthogonal group or the spin group. Furthermore from the beginning of the discovery of supersymmetry, the space $\mathfrak{t}_1$ was supposed to be spanned by the so-called *fermionic* or *spin charges*. Essentially what this means is that $\mathfrak{t}_1$ must have a *spin structure*, or as we shall see presently, spinorial. Because the spin modules are not modules for the orthogonal groups but their two-fold covrs, namely the spin groups, we must assume that L_0 is the spin group (although in the physical literature the groups never come in). The source of this restriction on $\mathfrak{t}_1$ can be traced to the fact that in quantum field theory the objects obeying the anticommutation rules are the spinor fields. To explain how special these groups are, even within the category of super semidirect products, let me go back to the 1970's. The first example of a super Lie algebra, the minimal supersymmetric extension of the Poincaré algebra, emerged at that time from the works of the Russian school, followed a little later by the construction, by Wess and Zumino, of the first

example of a simple super Lie algebra, and the general definition of super Lie algebras and the classification of the simple ones by Kac (see [10, 11, 12, 13] of previous essay). The initial idea was to construct a super Lie algebra

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1$$

with the following properties:

(a) $\mathfrak{g}$ is a *real* Lie super algebra and $\mathfrak{g}_0$ is the Poincaré algebra of 4-dimensional Minkowski spacetime.
(b) $\mathfrak{g}_1$ is a spinorial $\mathfrak{g}_0$-module which is irreducible over $\mathbf{R}$.
(c) The spacetime momenta should be among the super commutators $[A, B]$ where $A, B \in \mathfrak{g}_1$, i.e., $[\mathfrak{g}_1, \mathfrak{g}_1]$ should contain the translation subspace $\mathfrak{t}_0$ of $\mathfrak{g}_0$.
(d) $\mathfrak{g}$ should be minimal in the sense that there is no super Lie algebra $\mathfrak{h}$ such that $\mathfrak{g}_0 \subset \mathfrak{h} \subset \mathfrak{g}$ where both inclusions are proper.

The condition (b) includes the requirement that $\mathfrak{g}_0$ acts on $\mathfrak{t}_1$ through $\mathfrak{l}_0$ so that $\mathfrak{t}_0$ acts as zero. The minimality in (d) forces that $\mathfrak{t}_1$ is irreducible as a module for $\mathfrak{l}_0$ (over the real field); indeed, if $\mathfrak{t}_2$ is a submodule, $\mathfrak{g}_0 \oplus \mathfrak{t}_2$ is a super Lie algebra satisfying (a)-(c). The complexification of $\mathfrak{l}_0$ is a direct sum of two copies of $\mathfrak{sl}(2, \mathbf{C})$ and so the irreducible representations of $\mathfrak{l}_0$ are of the form $\mathbf{k} \otimes \mathbf{k}^{\text{conj}}$ where $\mathbf{k}$ is the irreducible representation of $\mathfrak{sl}(2, \mathbf{C})$ of dimension k. Spinoriality means that the complexification of $\mathfrak{t}_1$ must be a direct sum of the $\mathbf{2}$ and $\mathbf{2}^{\text{conj}}$. Now $\mathbf{2} \oplus \mathbf{2}^{\text{conj}}$ has a real form which is therefore the unique irreducible real spinorial module for $\mathfrak{l}_0$; it is called the *Majorana spinor* by the physicists. There is a unique (up to a multiplicative scalar) symmetric L_0-equivariant map $\mathfrak{t}_1 \otimes \mathfrak{t}_1 \longrightarrow \mathfrak{t}_0$ which then can be used for the bracket of the elements of $\mathfrak{t}_1$. This is the construction of the super Poincaré algebra.

The physicists take an explicit realization of $\mathbf{2}$ and the odd bracket, and write down the commutation rules for $\mathfrak{g}$. It is more convenient for me to write these in the following basis-independent form. Let the elements of $\mathbf{C}^2$ be viewed as column vectors and let

$$\mathfrak{t}_1 = \{(u, \bar{u}) | u \in \mathbf{C}^2\} \qquad \mathfrak{t}_{1\mathbf{C}} = \{(u, v) | u, v \in \mathbf{C}^2\}$$

with $\mathrm{SL}(2, \mathbf{C})$-actions given by

$$g, (u, \bar{u}) \longmapsto (gu, \bar{g}\bar{u}) \qquad g, (u, v) \longmapsto (gu, \bar{g}v).$$

Then $\mathfrak{t}_1$ is a real form of $\mathfrak{t}_{1\mathbf{C}}$ stable under this action, while $\mathfrak{t}_{1\mathbf{C}}$ is $\mathbf{2} \oplus \bar{\mathbf{2}}$. We take $\mathfrak{t}_0$ to be the space of 2×2 hermitian matrices represented as

$$x = \begin{pmatrix} x_0 + x_3 & x_1 + ix_2 \\ x_1 - ix_2 & x_0 - x_3 \end{pmatrix}, \quad \det(x) = x_0^2 - x_1^2 - x_2^2 - x_3^2$$

with L_0 and its action on $\mathfrak{t}_0$ given by

$$L_0 = \mathrm{SL}(2, \mathbf{C}) \qquad g, x \longmapsto gx\bar{g}^T.$$

Here the exponent T refers to the transposed matrix. The L_0-equivariant symmetric form $\mathfrak{t}_1 \times \mathfrak{t}_1 \longrightarrow \mathfrak{t}_0$ is then of the form

$$[(u, \bar{u}), (v, \bar{v})] = \frac{1}{2}(u\bar{v}^T + v\bar{u}^T).$$

Actually we can have any real constant in place of $1/2$ since only its sign matters for the isomorphism class of $\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{t}_1$ where $\mathfrak{g}_0 = \mathfrak{t}_0 \oplus \mathfrak{l}_0$. We take the factor to be $1/2$ (the negative sign gives a super Lie algebra which may be called an *isomer* of the one we use). This gives the super Poincaré algebra used in physics. The choice of the positive sign implies that the super bracket has a very nice *positivity property*. To formulate it let us first note that the forward light cone C^+ in $\mathfrak{t}_0$ is given in the usual coordinates as $x_0 > 0, x_0^2 - x_1^2 - x_2^2 - x_3^2 \geq 0$ which reduces to $\mathrm{Tr}(x) > 0, \det(x) \geq 0$ or equivalently, that

$$C^+ = \{x \in \mathfrak{t}_0 | x \neq 0 \text{ is nonnegative definite}\}.$$

We have

(supercommutator) $$[X, X] = u\bar{u}^T \qquad X = (u, \bar{u}) \in \mathfrak{t}_1$$

so that

$$[X, X] \in C^+ \text{ for all } 0 \neq X \in \mathfrak{t}_1.$$

Since any $X \in \mathfrak{t}_0$ which is nonnegeative definite can be written as $u\bar{u}^T$ for some $u \in \mathbf{C}^2$, it follows that

$$X \longmapsto [X, X]$$

maps $\mathfrak{t}_1 \setminus \{0\}$ *onto* C^+.

Let P_μ be the standard basis of Minkowski space. In the matrix identification, $P_0 = I$, the identity matrix. If e_1, e_2 are the standard basis vectors of $\mathbf{C}^2$, and $X_i = (e_i, e_i)$, then

$$P_0 = [X_1, X_1] + [X_2, X_2].$$

In any representation, P_0 gets mapped to $\rho(X_1)^2 + \rho(X_2)^2$ which is ≥ 0. This is the *positivity of the energy*.

If θ_α are the coordinates with respect to a basis for $\mathfrak{t}_1$, then x_μ, θ_α are the coordinates for the super spacetime $\mathfrak{t}$. The superfields, namely the sections of the structure sheaf $\mathcal{O}_{\mathfrak{t}_1}$, are thus of the form

$$\Phi(x, \theta) = s_0(x) + \sum_\alpha s_\alpha \theta_\alpha + \sum_{\alpha<\beta} s_{\alpha\beta}\theta_\alpha\theta_\beta + \sum_{\alpha<\beta<\gamma} s_{\alpha\beta\gamma}\theta_\alpha\theta_\beta\theta_\gamma + s_{1234}\theta_1\theta_2\theta_3\theta_4.$$

where the s's are functions of the x's. Using the differential and integral calculus on super manifolds we can then introduce super Lagrangians and do superfield theory.

It is in this framework, and on this 4|4-dimensional spacetime, that Salam and Strathdee (see [4a, 4b] of previous essay) defined the superfields and analyzed their transformation properties under the super Poincaré group and classified the super particles. In the physics literature, the technique of working with super spacetimes is known as the *method of superfields*.

The construction of super spacetimes and super Poincaré groups can be extended to the case when $\mathfrak{t}_0$ has arbitrary dimension D and Minkowski signature, and $\mathfrak{t}_1$ is not necessarily irreducible. The key is to show that there are always nontrivial L_0-equivariant symmetric bilinear maps

$$\mathfrak{t}_1 \times \mathfrak{t}_1 \longrightarrow \mathfrak{t}_0,$$

and indeed, which are in addition *positive* in the sense described above. One can classify all the symmetric equivariant maps; the positive ones form a nonempty open *cone*. This leads to a classification of all super Poincaré algebras. We shall not go into this classification here. See [7], pp. 90–94, 258–262, as also [8]. If N is the number of irreducible spin modules in $\mathfrak{g}_1$ physicists speak of N-extended super symmetry. The super Poincaré groups

are then the semi direct products of the classical group L_0 with the super spacetimes.

For deeper insights into supersymmetric physics one has to have a theory of *super homogeneous spaces*. For the construction of the general super homogeneous space, see [14]. One can use this theory as exemplified in loc. cit. to construct super conformal spacetimes.

Spinors

I have passed over in silence the highly nontrivial theory of *spinors* on which the construction of *all* super Poincaré algebras in arbitrary dimension and (Minkowskian) signature depend. However a few brief comments may be in order. The spin groups are the universal covering groups of the orthogonal groups, and the spin representations are certain special representations of the spin groups, determined by the Clifford algebra. The basic theory is over $\mathbf{C}$, but one needs the version of this theory over $\mathbf{R}$ so that it can be usefully applied to physical problems. For general references see [6, 7, 8, 9]. Physicists mostly ignore the spin groups and work infinitesimally; the Clifford algebra appears in their work through the yoga of the *Dirac gamma matrices*.

The spin representations were originally discovered by Elie Cartan when he classified the irreducible representations of the simple Lie algebras in 1913. For these the basic building blocks are the *fundamental representations* canonically associated to the nodes of their Dynkin diagrams. The spin representations are the ones for the orthogonal algebras associated to the right extreme nodes. For $\mathfrak{so}(2k)(k \geq 3)$ there are two of these, $S^{\pm}$, of dimension 2^{k-1}, while for $\mathfrak{so}(2k+1)$ there is just one, denoted by S, of dimension 2^k. For $\mathfrak{so}(4)$ which is only semi simple and $\simeq \mathfrak{so}(3) \oplus \mathfrak{so}(3)$, the diagram consists of two nodes which are not linked, and either of the two nodes may serve as the one defining the representation. Thus the two spin representations of $\mathfrak{so}(4)$ are the $\mathbf{2} \otimes 1$ and $1 \otimes \mathbf{2}$ encountered before. Cartan noticed that the spin modules do not occur in the decomposition of the *tensors* over the space on which the orthogonal groups are defined. Then in the 1930's when Dirac discovered his famous relativistic equations of the electron, he rediscovered these representations (in the special case of $\mathfrak{so}(4)$) and so $S^{\pm}$, S came to be known as *spin representations*; the elements of the spaces on which they act were then called *spinors*. Initially interest was centered only in the case of dimension 4 but this changed after Brauer and Weyl wrote their famous paper

on spinors in higher dimensions [6]. The higher dimensions arise in physics in models where spacetime is extended by multiplication with very small (!) compact Riemannian manifolds.

Let me now explain the connections with the notion of spin and the Dirac equation. When Dirac wanted to write a *first order* relativistic equation on flat Minkowski spacetime it was natural for him to look for it as a square root of the wave operator. He discovered that a square root does not exist over the complex numbers but does exist over the Clifford numbers (Clifford algebra). The first order operator with constant coefficients, called the *Dirac operator*, whose coefficients are in a representation (minimal) of the Clifford algebra, has therefore 4×4 matrices as coefficients, and so can operate only on 4-component wave functions, and these components describe the *spin* of the electron. This is the situation alluded to at the beginning of this section. However it was not enough to construct the Dirac operator over the Clifford algebra; it was also necessary to find out how it transformed under the Lorentz transformations. Dirac discovered that one has to go to the two-fold cover of the Lorentz group to write down the transformation law. The two-fold cover is naturally called the spin group, and its representation, the spin representation. It is coupled with the Clifford algebra right from the outset. In the modern theory, the spin group is imbedded inside the even part of the Clifford algebra.

The generalization of these ideas to arbitrary dimension offers no obstacle. But defining the Dirac operator over a *curved spacetime* turned out to be more subtle. The reason is that the spinor spaces associated to a quadratic vector space V *do not vary functorially with* V. The attempt to formulate the Dirac equations and spinor fields over arbitrary Riemannian and Minkowskian manifolds led to the discovery that there are global topological obstructions to defining them. On a Riemannian (resp. Minkowskian) manifold we always have the Laplacian (resp. the D'Alembertian, wave) operator. The Dirac operator, being a square root of these, needs a finer structure to make sense, namely, the *spin structure*. When the obstructions vanish, the manifold is called a *spin manifold*.

The definition of spin modules given at the beginning, although historically correct, is not satisfactory in the sense that it does not allow us to explore their structure in detail. To get to this we introduce the *Clifford algebra* $C(V) = C$ of the complex *quadratic* vector space V (which is just a vector space equipped with a nondegeberate quadratic form). The algebra C is a *super algebra* and the Lie algebra $\mathfrak{g}$ of the orthogonal group has a natural imbedding inside the even part C_0 of C whose image generates C_0. Thus C_0 is *an enveloping*

algebra of $\mathfrak{g}$. The spin modules can then be proved to be precisely those modules for $\mathfrak{g}$ which extend to C_0; the proof is nontrivial, see Chapter 5 of [7], and [8]. Thus the spin modules can be identified with C_0-modules, the so-called *Clifford modules*. It is this identification that leads to the possibility of a very detailed study of spin modules. If (e_i) is an orthonormal basis for V, the Clifford algebra C is generated by the e_i with relations

$$e_i e_j + e_j e_i = 2\delta_{ij} \quad (1 \leq i, j \leq n).$$

In any representation of C the e_i are represented by matrices γ_i satisfying the commutation rules

$$\gamma_i \gamma_j + \gamma_j \gamma_i = 2\delta_{ij} \quad (1 \leq i, j \leq n)$$

which are called (after Dirac's pioneering constructions and notations) the *Dirac gamma matrices*. The yoga of gamma matrices is nothing but a coordinate version of the theory of Clifford algebras.

The spin representations are defined over the complex numbers. In physics one works over the real field and it becomes important to know what happens then. As we have done above we call a real representation of the orthogonal Lie algebra *spinorial* if it complexifies to a direct sum of spin modules. For instance the space $\mathfrak{t}_1$ in the super Poincaré Lie algebra is spinorial as it complexifies to $\mathbf{2} \oplus \bar{\mathbf{2}}$. Now if we have a finite dimensional irreducible real representation of some real associative algebra, its commutant is a division algebra over $\mathbf{R}$ and so is one of $\mathbf{R}$, $\mathbf{C}$, $\mathbf{H}$ by a classical theorem due to Frobenius; here $\mathbf{C}$ is treated as a real algebra, and $\mathbf{H}$ is the algebra of quaternions. Accordingly, the real representation is said to be of *type* $\mathbf{R}$, $\mathbf{C}$, $\mathbf{H}$ respectively. In the case above, the Majorana spinor is of type $\mathbf{C}$ because its commutant, when complexified, is the commutant of a direct sum of two *inequivalent* representations, and hence has dimension 2. It turns out that for the general case of V of dimension of signature s, the types of the *real* irreducible spinorial modules depend *only on the signature* s *mod* 8 and are given by the following table (see [7, 8] for a derivation of these tables).

Space of arbitrary signature s

s mod 8	type
0	**R**, **R**
1, 7	**R**
2, 6	**C**
3, 5	**H**
4	**H**, **H**

For Minkowski signature in dimension D, we have $s = -(D-2)$, and the above table becomes the following.

Minkowski space of dimension D

D mod 8	type
0, 4	**C**
1, 3	**R**
2	**R**, **R**
5, 7	**H**
6	**H**, **H**

One is also interested in supersymmetric Lagrangians and mass terms. These depend on the structure of real invariant forms and maps. There is a corresponding table but it is now determined by *both* dimension and signature mod 8. The physicists rely on the famous *Regge clock* whose two hands describe the dimension and signature mod 8. See [7, 8] for the full theory (Chapters 5 and 6) and references. The periodicity mod 8 is a real marvel of this theory. For a theory of spinors over *any* field, see Chevalley's great treatise [9]. For the Regge clock see [10]. For a general treatment of physics and geometry, see [15].

5.5 Unitary representations of super Poincaré groups

The theory discussed above applies at once for $S = (G_0, \mathfrak{g})$ a super Poincaré group. Because of the Minkowski form we can identify $\mathfrak{t}_0^*$ with $\mathfrak{t}_0$ itself. For $p \in \Gamma^+$ the stabilizer L_0^p is given by

$$L_0^p = \begin{cases} \mathrm{Spin}(1, D-1) & \text{if } \langle p, p\rangle > 0 \\ \mathbf{R}^{D-2} \times' \mathrm{Spin}(D-2) & \text{if } \langle p, p\rangle = 0,\ p_0 > 0 \\ L_0 & \text{if } p = 0. \end{cases}$$

Thus, except when $D = 4$, $p_0 > 0$, $\langle p, p\rangle = 0$, the stabilizer L_0^p is simply connected; in the exceptional case of zero mass in 4 space time dimensions, the stabilizer is $\mathbf{R}^2 \times' T$ where T is a circle group and so still falls within the condition (**C**) described in Section 5.2.

The key lemma which shows the nature of T^+ in this case is the following.

Lemma 2. *We have*

$$Q_p \geq 0 \Longleftrightarrow p_0 \geq 0, \langle p, p\rangle \geq 0.$$

Proof (Sketch). If $v \in (\Gamma^+)^0$, $\langle v, [X, X]\rangle > 0$ for all $0 \neq X \in \mathfrak{g}_1$ by the structure of the super Poincaré algebra (positivity of the odd commutators). By going to the limit we see that for $p \in \Gamma^+$, $\langle p, [X, X]\rangle \geq 0$ for all $X \in \mathfrak{g}_1$. This is just the statement that $Q_p \geq 0$. For the converse, suppose that $Q_p \geq 0$ but $\langle p, p\rangle < 0$. Then $L_0^p \simeq \mathrm{Spin}(1, D-2)$. On the other hand we have a map $L_0^p \longrightarrow \mathrm{SO}(\mathfrak{g}_{1p})$. Since $\mathrm{Spin}(1, D-2)$ is simple and non compact it cannot have a nontrivial map into a compact group. Hence L_0^p acts trivially on $\mathfrak{g}_{1p}$. Thus the action of L_0^p on $\mathfrak{g}_1$ contains the trivial representation of L_0^p. But L_0 acts as a spin module and it is not difficult to show that the action of L_0^p is a sum of spin modules, and so cannot contain the trivial representation.

Theorem 9. *The irreducible unitary representations of a super Poincaré group $S = (G_0, \mathfrak{g})$ are parametrized by the orbits of p with $p_0 \geq 0$, $\langle p, p\rangle \geq 0$, and for such p, by irreducible unitary representations of the stabilizer L_0^p at p. Let τ_p be an irreducible SA representation of the Clifford algebra $\mathcal{C}_p$ and let κ_p be the representation of L_0^p in the space of τ_p defined earlier. Then, for any irreducible unitary representation θ of L_0^p the pair (σ, ρ^σ) defined by*

$$\sigma = e^{ip}\theta \otimes \kappa_p, \quad \rho^\sigma(X) = 1 \otimes \tau_p(X) \quad (X \in \mathfrak{g}_1)$$

is an irreducible unitary representation of the little super group $S^p = (T_0 L_0^p, \mathfrak{t}_0 \oplus \mathfrak{l}_0^p \oplus \mathfrak{g}_1)$, and all irreducible unitary representations of S^p are obtained in this manner. The unitary representation $\ominus_{p\theta}$ of the super Lie group S induced by it is irreducible and all irreducible unitary representations of S are obtained in this manner, the correspondence $(p, \theta) \longmapsto \ominus_{p\theta}$ being bijective up to equivalence.

Remark 3. The reader can see that in the presence of supersymmetry, only representations of positive energy come in, in contrast to the classical situation where we could not exclude the tachyonic orbits. Otherwise the reduction to the little group is exactly as before, parametrized by the irreducible unitary representations of L_0^p, except that it has to be tensored with the pair $(\kappa_p, \tau_p,)$. Thus the determination of (κ_p, τ_p) is the basic new ingredient.

Remark 4. The correspondence with representations of the little group allows us to treat zero mass super particles with infinite spin, somehting not available in the physics literature. Also the SUSY transformations are given on the full space of super particle states, not only on those with a fixed momentum.

5.6 Super particles and their multiplet structure

The irreducible unitary representations of the super Poincaré group are the models for *super particles.* If we restrict the representation of G_0, it will in general decompose into a direct sum of particles and this set is called the *multiplet* defined by the super particle. The operators $\rho(X)(X \in \mathfrak{g}_1)$ then give the transformations between particles of different spin parities. The fundamental muptiplet is the one corresponding to the trivial representation of L_0^p. The irreducible unitary representations into which κ_p splits then define the classical particles of the fundamental multiplet.

The calculation of κ_p can be carried out very explicitly. When $D = 4$ and $\mathfrak{g}_1$ is the Majorana spinor, the multiplet with mass $m > 0$ has the spins

$$\begin{cases} \{j, j, j + 1/2, j - 1/2 (j > 0) \\ \{0, 0, 1/2\} (j = 0). \end{cases}$$

For $m = 0$ the mulitplet has spins (helicities)

$$\{n/2, (n + 1)/2\}.$$

These results are the source for the prediction that supersymmetry forces the known elementary particles to have *super partners.* Whether these partners can be seen when the new LHC opens is at best speculative, since we do not know the threshold of supersymmetry breaking in the development of the early universe.

References

[1a] C. Carmeli, G. Cassinelli, A. Toigo, and V. S. Varadarajan, *Unitary representations of super Lie groups and applications to the classification and multiplet structure of super particles*, Comm. Math. Phys., 263(2006) 217–258.

[1b] V. S. Varadarajan, *Unitary representations of super Lie groups and some applications*, The Math. Student Special Centenary Volume, (2007), 33–57.

[2] C. Fronsdal, and T. Hirai, *Unitary representations of supergroups*, 15–66, in *Essays on Supersymmetry, with contributions by* M. Flato, C. Fronsdal, and T. Hirai, Mathematical Physics Studies, Vol. 8, 1986, D. Reidel, C. Fronsdal (Ed.).

[3] E. P. Wigner, *Unitary representations of the inhomogeneous Lorentz group*, Ann. Math., 40(1939), 149–204.

[4] For a wide-ranging retrospective on Mackey and his work, see the following.

V. S. Varadarajan, *George Mackey and his work on representation theory and foundations of physics*, 417–446, in *Group Representations, Ergodic Theory, and Mathematical Physics : A Tribute to George Mackey*, Contemp. Math., 449, AMS., 2008, R. S. Doran, C. C. Moore, R. J. Zimmer, (Eds.).

[5] E. G. Effros, *Transformation Groups and C^*-algebras*, Ann. Math., 81(1965), 38–55.

[6] R. Brauer, and H. Weyl, *Spinors in n dimensions*, American J. Math., 57(1935), 425–449; also H. Weyl, *Gesammelte Abhandlungen*, vol. III, 493–516, Springer, Berlin-New York 1968.

[7] V. S. Varadarajan, *Supersymmetry for Mathematicians: An Introduction*, AMS-Courant Institute Lecture Notes, Providence, R.I., 2004.

[8] P. Deligne, *Notes on spinors*, in *Quantum Fields and Strings: a Course for Mathematicians*, vol. I, 99–135, American Mathematical Society, Providence, R.I.,1999.

[9] C. Chevalley, *The Algebraic Theory of Spinors and Clifford Algebras*. Collected Works. Vol. 2. P. Cartier and C. Chevalley, eds. Springer, Berlin, 1997. This is a reprint of the original book written in 1954.

[10] T. Regge, *The group manifold approach to unified gravity*, in *Relativity, groups, and topology II*(Les Houches, 1983), 933-1005. North-Holland, Amsterdam, 1984.

[11] H. P. Jakobsen, *The full set of unitarizable highest weight modules of basic classical Lie super algebras*, Memoirs of the Amer. Math. Soc., 532 (1994), Providence, R. I.

[12] Dobrev, V. K., and Petkova, V. B., *All positive energy unitary irreducible representations of extended conformal super symmetry*, Phys. Lett. B 162(1-3), 127-132 (1985); *All positive energy unitary irreducible representations of the extended conformal super algebra*. In *Conformal groups and related symmetries: physical results and mathematical background*, Clausthal-Zellerfeld, 1985, Vol 261, Lecture Notes in Physics, Springer, Berlin, 1986, 300-308; *Group theoretical approach to extended conformal super symmetry: function space realization and invariant differential operators*, Fortschr. Phys. 35 (7), 537-572 (1987).

[13] H. Salmasian, *Unitary representations of nilpotent super Lie groups*, Comm. Math. Phys., 297(2010), 189–227.

[14] R. Fioresi, M. A. Lledó, and V. S. Varadarajan. *The Minkowski and conformal superspaces*, J. Math. Phys., 48(11)113505, 27, 2007.

[15] E. Witten, *Physics and geometry*, in Proceedings of the Int. Congress of Mathematicians, Vol 1, 267–303, Berkeley, 1986-1987.

Chapter 6
Nonarchimedean physics*

> We will argue that for a physical theory of quanta there are only two new possibilities, namely the field of p-adic numbers or a finite Galois field.
> *Igor Volovich*

There is a natural time scale, the *Planck scale*, that emerges when general relativity and quantum mechanics are both significant. No measurements are possible in regions of smaller sizes than the Planck units. In particular we cannot compare distances and times in sub-Planckian domains. In 1987 Igor Volovich proposed the bold hypothesis that in such domains the geometry of spacetime is *nonarchimedean*. It is then natural to ask what are the consequences of this hypothesis. We address this question in the Dirac mode.

6.1. Planck scale and the Volovich hypothesis
6.2. Nonarchimedean symmetry groups and their multipliers
6.3. Elementary particles over nonarchimedean fields
6.4. Nonarchimedean quantum field theory

6.1 Planck scale and the Volovich hypothesis

That the structure of spacetime itself could be the subject of investigation is an idea that dates back to the 1970's, when many physicists, concerned about the divergences in quantum field theories, started exploring the micro-structure of spacetime as a possible source of these problems. In particular, Beltrametti and his collaborators proposed the idea in [1, 2, 3] that the geometry of spacetime might be based on a nonarchimedean, or even a finite, field and examined some of the consequences of this hypothesis. But the idea did not really take off until Volovich [4, 5] proposed in 1987 that world geometry

* This essay is based on various lectures I have given for the past several years. I especially want to mention conferences in Goslar, in Dubna, in Moscow, and a series of lectures in the University of Arizona, sponsored by my friend Hermann Flaschka.

in regimes smaller than the Planck units might be nonarchimedean. The reasoning behind this hypothesis is this: since no measurements are possible at such ultra-small distances and time scales due to the interplay between general relativity and quantum theory, comparisons of distances and time spans are also ruled out.

The idea of a fundamental scale emerges naturally when one attempts to unify gravitation and quantum theory. This scale, called the *Planck scale*, is one below which it seems that no measurements can be carried out. The idea of *extended particles* (*strings, branes*) also makes sense in this context. Indeed, the Planck scale comes up immediately when one identifies the Schwarzchild radius of a particle, basic in general relativity, with its Compton wave length, fundamental in quantum theory. This leads to the order of Planck length as 10^{-33} cm and Planck time (time taken by light to travel Planck length) as 10^{-43} sec.

The Planck scale actually goes back all the way to Planck himself. He wanted to obtain a fundamental scale entirely in terms of the 3 constants $\hbar$, G, c where G is the gravitational constant. We have

$$\begin{aligned}\hbar &\sim 6 \times 10^{-27}\ cm^2\ gm\ sec^{-1}\\ G &\sim 6.7 \times 10^{-8}\ cm^3\ gm^{-1}\ sec^{-2}\\ c &\sim 3 \times 10^{10}\ cm\ sec^{-1}.\end{aligned}$$

It follows after a simple calculation that we have the following quantities which have the dimensions of *length*, *mass*, and *time*, the so-called *Planck units*, which form the *Planck scale*:

$$\begin{aligned}\ell_{\mathrm{Pl}} &= \left(\frac{\hbar G}{c^3}\right)^{1/2} \sim 10^{-33}\ cm\\ M_{\mathrm{Pl}} &= \left(\frac{\hbar c}{G}\right)^{1/2} \sim 10^{-4}\ gm\\ t_{\mathrm{Pl}} &= \left(\frac{\hbar G}{c^5}\right)^{1/2} \sim 10^{-43}\ sec.\end{aligned}$$

From this point of view, a *Planck particle* is a black hole whose event horizon has a radius equal to its Compton wave length, as we described earlier.

Since impossibility of measurements automatically forbids also comparisons between different lengths and also different times, the Volovich hypothesis is very natural. Now nonarchimedean geometries arise when we work

over the field $\mathbf{Q}_p$ of p-adic numbers, p being a rational prime ≥ 2. Since no single prime can be given distinguished status, it is perhaps even more natural to see if one could really work with an *adelic* geometry as the basis for space-time. Such an idea was first proposed by Manin [6]. A huge number of articles have appeared since then, exploring these and related themes. For a definitive survey and a very inclusive set of references see the very recent article by Dragovich et al [7].

The simplest way to enforce nonarchimedean geometry is to assume that spacetime is a manifold over $\mathbf{Q}_p$. Over the reals Wigner associated elementary particles with irreducible projective unitary representations of the Poincaré group, and classified these in his epoch-making paper of 1939. In the paradigm of modern quantum field theory one goes beyond the Wigner classification and defines elementary particles as the quanta of quantum fields [8]. However, over the p-adic numbers it has not yet been found possible to develop much intuition about p-adic quantum fields that are fully covariant, for many reasons, such as the absence of a Lagrangian picture. So one has to be content with thinking of elementary particles as defined by irreducible projective unitary representations of the symmetry groups, the p-adic analogues of the real Poincaré and Galilean groups. This means also that one has to define the spacetime symmetry groups as p-adic Lie groups and study their projective representations.

There are many studies designed to show that at Planckian distances and times spacetime itself undergoes fluctuations. It is quite plausible that such fluctuations change the ground field from $\mathbf{R}$ to the $\mathbf{Q}_p$. The themes discussed in this essay are thus of interest, at least in the Dirac mode whose essence is: do the mathematics first and then look for physical interpretation.

6.2 Nonarchimedean symmetry groups and their multipliers

How do we formulate the notion that a quantum system has a symmetry group G? A symmetry of a quantum system is a bijection of the projective space $P(\mathcal{H})$ of the Hilbert space $\mathcal{H}$ of the system preserving the transition probability $|(\phi, \psi)|^2$ between two states $[\phi]$, $[\psi]$ defined by unit vectors ϕ and ψ. It is a classical theorem of Wigner and Eckhart that if σ is a symmetry, then σ is of the form

$$[\phi] \longmapsto [R\phi]$$

where $R = R_\sigma$ is a unitary or anti-unitary operator; R is uniquely determined by the symmetry up to a phase factor. So, if $\sigma, \sigma_1, \sigma_2$ are symmetries with $\sigma = \sigma_1\sigma_2$, and R, R_1, R_2 are operators inducing them, we cannot be sure that $R = R_1R_2$ but only that $R = mR_1R_2$ where m is a complex number of absolute value 1. It follows from this that the symmetry of a quantum system with respect to a group G, locally compact and second countable (lcsc), may be expressed by a projective unitary representation (PUR), either of G or of a subgroup of index 2 in G, in the Hilbert space of quantum states. Here we must recall that a PUR, to be thought of as a map of the group into the projective unitary group of a Hilbert space, may be viewed as a Borel map $x \longmapsto U(x)$ of G into the unitary group of $\mathcal{H}$ such that

$$U(x)U(y) = m(x, y)U(xy), \qquad U(1) = 1 \quad (x, y \in G).$$

The function m, with values in the circle group, is called a *multiplier* for U, and U is called an *m-representation*. It satisfies certain identities coming from the associativity of multiplication and we call any such function a multiplier; the point is that every such function arises as the multiplier of some projective representation. The relations satisfied by m are

$$m(xy, z)m(x, y) = m(x, yz)m(y, z), \qquad m(x, 1) = m(1, x) = 1.$$

The projective representation does not change if we make the change

$$U(x) \longmapsto a(x)U(x);$$

but in this case, m changes to m' where

$$m'(x, y) = a(x)a(y)a(xy)^{-1}m(x, y).$$

We write $m \sim m'$. The multipliers form an abelian group under point-wise multiplication, those ~ 1 form a subgroup, and the quotient is the *multiplier group* $H^2(G)$ of G. This is an important (cohomological) invariant of G and its determination is basic to all investigations.

The insistence on *Borel* multipliers is not out of a desire for generality. On general grounds projective representations of a group G are the same as ordinary representations of a central extension E of G by a circle group T. However the map $E \to G$ may not have continuous sections. It was

Mackey who discovered that Borel sections always exist giving rise to Borel multipliers.

If U is an m-representation we introduce the group $E_m = T \times G$, T being the group of complex numbers of modulus 1, with the multiplication law

$$(s, x)(t, y) = (stm(x, y), xy).$$

It is easy to verify that the properties of m guarantee that E_m is a group, with natural maps

$$T \longrightarrow E_m, \quad t \longmapsto (t, 1) \qquad E_m \longrightarrow G, \quad (s, x) \longmapsto x.$$

At the level of abstract groups we have a central extension

$$1 \longrightarrow T \longrightarrow E_m \longrightarrow G \longrightarrow 1.$$

Then

$$V : (s, x) \longmapsto sU(x)$$

is an ordinary representation of E_m, giving us a way to remove the multiplier by passing to a central extension of G. But the topologies are missing; we have not given a recipe for viewing E_m as a topological group, much less as an lcsc group. If m is continuous, then we can use the product topology for E_m; otherwise we use a beautiful theorem of Mackey that there is a *unique* topology for E_m in which it is lcsc and the above is a central extension of lcsc groups. The topology is the *Weil topology*; we consider the Hilbert space $L^2(E_m)$ with respect to the measure $dtdx$ where dt is the usual Lebesgue measure on T and dx is a left Haar measures on G, and give the topology for E_m in which the left regular representation of E_m is a homeomorphism into the unitary group of $L^2(E_m)$ with its strong operator topology [9].

The central extension E_m depends on m and so it is reasonable to ask if we can find a *universal* central extension G^* of G with the property that all PURs of G are unitarizable when lifted to G^*. This is not always possible; the commutator subgroup of G must be dense in G for a universal topological central extension (UTCE) to exist. For the usual Poincaré group $P = R^4 \times' \mathrm{SO}(1, 3)^0$, E. Wigner [9, 10], proved that all PURs of the connected *real* Poincaré group P lift to URs of the simply connected covering group $P^* = \mathbf{R}^{1,3} \times' \mathrm{Spin}(\mathbf{R}^{1,3})$ of P where $\times'$ denotes semidirect product. In other words, P^* is already the UTCE of the Poincaré group.

Central extensions go back at least to Schur who calculated them for the symmetric and alternating group. Weyl [11] was the first to emphasize the importance of PURs in quantum theory. He called them *ray representations*. Bargmann [12] extended Wigner's work to all dimensions, besides giving a systematic treatment of multipliers of Lie groups and treating the case of the Galilean groups also. For a deep and systematic treatment of the Galilean group see [13]. For varied aspects of the theory of multipliers see [14, 15, 16, 17, 18a, 18b].

The situation over a disconnected field is quite different. Over a nonarchimedean local field k , let G be an algebraic group defined over k and $G(k)$ the lcsc group of its k-points. The commutator subgroups of $G(k)$ are often open and closed subgroups of $G(k)$, and it is generally a very delicate matter to verify whether they are equal to $G(k)$. So it is preferable to work with the PURs of $G(k)$ itself, rather than look for UTCEs of $G(k)$. Number theorists have long been interested in URs of groups $G(k)$ for *simple* groups G. In physics, groups with radical appear to be important and so it is worthwhile to study PURs of these groups as well. The groups $G(k)$ are all totally disconnected. For totally disconnected groups David Wigner proved that all multipliers are equivalent to continuous ones [14].

Let $k = \mathbf{Q}_p$. Actually most of what I say will remain valid if the characteristic of k is $\neq 2$. Let me introduce the Poincaré group. Let $\mathbf{V}$ be an isotropic quadratic vector space over k; this means that $\mathbf{V}$ has a nondegenerate quadratic form defined over k which has null vectors over k. Then we have the algebraic groups $\mathbf{G} = \mathrm{SO}(\mathbf{V})$ and its two-fold algebraic cover $\mathbf{G}_{spin} = \mathrm{Spin}(\mathbf{V})$. We thus have correspondingly the Poincaré groups $\mathbf{P} = \mathbf{V} \times' \mathbf{G}$ and $\mathbf{P}_{spin} = \mathbf{V} \times' \mathbf{G}_{spin}$. Write V, G and G_{spin} for the groups of k-points of $\mathbf{V}$, $\mathbf{G}$, and $\mathbf{G}_{spin}$ respectively, and P, P_{spin} for the respective groups of k-points of $\mathbf{P}$, $\mathbf{P}_{spin}$. Unlike the case when $k = \mathbf{R}$, we can no longer speak of the signature but the *Witt class* of V.

Moore [15] first calculated the UTCEs of simple groups split over k (Chevalley groups), and then Prasad and Raghunathan [16] extended his results (amd much more) for nonsplit but isotropic groups. Now G_{spin} and P_{spin} *do have* TCE's; for the spin groups this is a consequence of the work of Moore and Prasad and Raghunathan, and for the corresponding Poincaré groups, of the work of Varadarajan [17]. Moreover, if G^*_{spin} is the universal TCE of G_{spin}, it is shown in [17] that the universal TCE P^*_{spin} of P_{spin} is given by $P^*_{spin} = V \times' G^*_{spin}$. So all PURs of P_{spin} lift to URs of P^*_{spin}, and since P^* is a semidirect product, the Mackey-Wigner theory is applicable. We are thus

in the same situation as in the real case and there are no fundamental obstacles to the classification of the particles (=irreducible PURs) with P_{spin}-symmetry.

However the natural maps $G_{spin} \longrightarrow G$ and $P_{spin} \longrightarrow P$ are *not* surjective (even though they are surjective over the algebraic closure of k), and so replacing the orthogonal group G by the spin group G_{spin} leads to a loss of information. So we work with the orthogonal group rather than the spin group. To illustrate this point, let $G = \mathrm{SL}(2, \mathbf{Q}_p)$. The adjoint representation exhibits G as the spin group corresponding to the quadratic vector space $\mathfrak{g}$ which is the Lie algebra of G equipped with the Killing form. The adjoint map $G \longrightarrow G_1 = \mathrm{SO}(\mathfrak{g})$ is the spin covering for $\mathrm{SO}(\mathfrak{g})$ but this is *not surjective*; in the standard basis

$$X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

the spin covering map is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto \begin{pmatrix} a^2 & -2ab & -b^2 \\ -ac & ad+bc & bd \\ -c^2 & 2cd & d^2 \end{pmatrix}.$$

The matrix

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha^{-1} \end{pmatrix}$$

is in $\mathrm{SO}(\mathfrak{g})$; if it is the image of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $b = c = 0$ and $\alpha = a^2$, so that unless $\alpha \in \mathbf{Q}_p^{\times 2}$, this will not happen.

The group P is still a semidirect product but we are now required to determine its irreducible *projective* URs. This means that we must determine its multipliers and then, to each multiplier m, find the irreducible m-representations. The multipliers of a semidirect product were determined by Mackey long ago [17, 18]. It turns out further that there is a very nice variant of the Mackey machine for m-representations of a seimidirect product that allows us to do this [19], which appears not to have been noticed in previous literature. The variant is formulated in the framework of locally compact groups and so is applicable to adelic geometries as well, but here we restrict ourselves to the local case.

Mackey's formula for multipliers of semidirect products

I shall now describe Mackey's beautiful formula for multipliers of semidirect products. Let $H = A \times' G$ where A and G are lcsc groups and A is abelian. Let A^* be the character group of A. We need to introduce $H^1(G, A^*)$ which we do as usual. An 1-cocycle for G with coefficients in A^* is a Borel map $f(G \longrightarrow A^*)$ such that

$$f(gg') = f(g) + g[f(g')] \quad (g, g' \in G)$$

or equivalently that $g \mapsto (f(g), g)$ is a Borel homomorphism of G into the semidirect product $A^* \times' G$. Thus all 1-cocycles are continuous. The abelian group of continuous 1-cocycles is $Z^1(G, A^*)$ and the coboundaries are cocycles of the form $g \mapsto g[\chi] - \chi$ for some $\chi \in A^*$. These form a subgroup $B^1(G, A^*)$ of $Z^1(G, A^*)$ to give the cohomology group $H^1(G, A^*) = Z^1(G, A^*)/B^1(G, A^*)$.

In practice we are primarily interested not in the whole of $H^2(H)$ but only in $M_A(H)$, the group of multipliers on H that are trivial when restricted to $A \times A$. Here we must remember [11] that abelian groups can have nontrivial multiplier groups and so the above requirement is really a condition on the way that G operates on A; I shall comment on this presently. Let $H^2_A(H)$ denote the image of $M_A(H)$ in $H^2(H)$. Let $M'_A(H)$ be the group of multipliers m for H with $m|_{A\times A} = m|_{A\times G} = 1$. From [17, 18] we find the isomorphism

$$H^2_A(H) \simeq H^2(G) \times H^1(G, A^*).$$

More precisely, any element in $M_A(H)$ is equivalent to one in $M'_A(H)$. If $m \in M'_A(H)$, $m_0 = m|_{G\times G}$, and $\theta_m(g^{-1})(a') = m(g, a')$, then $\theta_m \in Z^1(G, A^*)$, and $m \mapsto (m_0, \theta_m)$ is an isomorphism $M'_A(H) \simeq Z^2(G) \times Z^1(G, A^*)$ which is well defined in cohomology and gives the above isomorphism. It follows from this that if $n \in Z^2(G)$ and $\theta \in Z^1(G, A^*)$ are given then one may define $m \in M'_A(H)$ by $m(ag, a'g') = n(g, g')\theta(g^{-1})(a')$. If $n = 1$ then $m(ag, a'g') = \theta(g^{-1})(a')$. If $H^1(G, A^*) = 0$ every multiplier of H is equivalent to the lift to H of a multiplier for G.

As I mentioned above, abelian groups in general have nontrivial multiplier groups. If μ is a multiplier for A and $g \in G$, then

$$\mu^g : a, b \longmapsto \mu(gag^{-1}, gbg^{-1})$$

is again a multiplier for A, and the map $\mu \longmapsto \mu^g$ induces a map

$$H^2(A) \longrightarrow H^2(A).$$

As g varies we obtain an action of G on $H^2(A)$. Suppose now that U is a PUR of H with multiplier m and let m_A be the restriction of m to $A \times A$. Then, for fixed $g \in G$, m_A is once again the multiplier of the PUR $a \longmapsto U(g)U(a)U(g)^{-1}$ of A. On the other hand, for some constant $c(a)$ with $|c(a)| = 1$ we have

$$U(g)U(a)U(g)^{-1} = c(a)U(gag^{-1}) \qquad (a \in A)$$

and so the representations

$$a \longmapsto U(g)U(a)U(g)^{-1} \quad \text{and} \quad a \longmapsto U(gag^{-1}) \qquad (a \in A)$$

are projectively the same. Since m_A^g is the multiplier of the latter PUR, it follows that m_A and m_A^g are equivalent. Hence the corresponding element in cohomology, the class $[m_A]$ of m_A, is fixed by g. Thus $[m_A]$ is G-invariant. If we know that G does not fix any nonzero element of $H^2(A)$, then m_A is trivial, and we are in the framework of Mackey's formula discussed above.

Let us write $\Lambda^2(A)$ for the group of alternating bicharacters of A, namely maps $f(A \times A \longrightarrow T$ which are characters in each variable separately, and satisfy $f(a, b)f(b, a) = 1$. It is then easy to check that such f are multipliers for A and we have a map $\Lambda^2(A) \longrightarrow H^2(A)$. If we now assume that A is 2-regular, i.e., the map $a \longmapsto 2a$ is an automorphism of A, it is then shown in [20] that the above map is an isomorphism, i.e.,

$$\Lambda^2(A) \simeq H^2(A).$$

The conditions

$$A \quad \text{is 2-regular,} \quad \Lambda^2(A)^G = 0$$

then guarantee that all multipliers of H are trivial on A.

For instance, if V is a finite dimensional vector space over $\mathbf{R}$ or $\mathbf{Q}_p$, then V is 2-regular and $\Lambda^2(V)$ is naturally isomorphic to the space of skew symmetric bilinear forms on $V \times V$, as was shown in [17]. If b is a skew symmetric bilinear form on $V \times V$, the map $m_b : x, y \longmapsto e^{ib(x,y)}$ is a multiplier for V, and $b \longmapsto m_b$ induces an isomorphism $\Lambda^2(V) \simeq H^2(V)$. If we have a PUR of H where $A = V$, then the bilinear form b has to be invariant under G. If

G leaves a symmetric nondegenerate bilinear form invariant and if the action of G is irreducible, then the space of invariant bilinear forms is of dimension 1, and so b must be symmetric, hence 0. In particular, for the Poincaré group this means that $b = 0$ and we are in the framework discussed in Mackey's result.

Mackey machine for PURs of semidirect products

If U is a UIR of H, its restriction to A can be described by a projection valued measure P on A^* by

$$U(a) = \int_{A^*} \chi(a) dP(\chi) \quad (a \in A);$$

to say that the spectrum of U is contained in a set E simply means that

$$P(A^* \setminus E) = 0.$$

This said, let us continue with $H = A \times' G$. The Mackey machine for ordinary representations operates as follows. We look at the action of G on A^*. Let O be any orbit for this action. To this orbit one associates a set of UIRs of H. We fix a $\chi \in O$ and let G_χ be the stabilizer of χ in G. Then to any UIR ρ of G_χ is associated a UIR $T_{O,\rho}$ of H; the equivalence class of $T_{O,\rho}$ depends only on that of ρ, and inequivalent ρ's give rise to inequivalent T's. The URs $T_{O,\rho}$ have their spectra contained in the orbit O. If we change to a different χ in O, the set of UIRs of H described by this procedure does not change. One can describe this set as *precisely* the set of all UIRs of H whose spectra are contained in O. Clearly, the T's associated to distinct O's are inequivalent, since their spectra are contained in different, hence disjoint, orbits. As we vary O we thus obtain a disjoint family of subsets of the unitary dual of H.

Under a regularity condition on the action of G on A^* this procedure, called the *Mackey machine*, gives *all* of the unitary dual of H. The condition is that the action of G on A^* is *Borel smooth*, i.e., there is a Borel set in A^* which meets each orbit of G exactly once. What happens is that in the presence of regularity, *the spectrum of any UIR of H is necessarily contained in some orbit*. It is a famous and beautiful theorem of Effros [21] that the action is regular if and only of each orbit is locally closed, i.e., open in its closure. Of

course the action need not always be regular. Non regular actions are often called *ergodic*. In the ergodic case there are UIRs of H whose spectra are *not* contained in any orbit, and so these cannot obtained by the Mackey machine [9]. The complete determination of the unitary dual of H in such cases is still elusive, as far as I know.

It turns out [19] that there is a remarkable variant of the Mackey machine for dealing with PUIRs of H. In order to describe this we shall now introduce certain new actions of G on A^* defined by cocycles in $Z(G, A^*)$. If $\theta : G \to A^*$ is a continuous map with $\theta(1) = 0$, then, defining $g\{\chi\} = g[\chi] + \theta(g)$, for $g \in G$, $\chi \in A^*$, it is easy to see that $g : \chi \mapsto g\{\chi\}$ defines an action of G on A^* if and only if $\theta \in Z^1(G, A^*)$. This action depends on the choice of the cocycle $\theta \in Z^1(G, A^*)$, so we write it as $g_\theta\{\chi\}$. The actions defined by θ and θ' are equivalent in the following sense: if $\theta'(g) = \theta(g) + g[\xi] - \xi$ where $\xi \in A^*$, then $g_{\theta'} = \tau^{-1} \circ g_\theta \circ \tau$ where τ is the translation by ξ in A^*. The action $g_\theta : \chi \mapsto g_\theta\{\chi\}$ is called *the affine action* of G on A^* determined by θ.

We now obtain the basic theorem of irreducible m-representations of H, which is a variant of the Mackey machine.

Theorem. *Fix $\chi \in A^*$, $m \simeq (m_0, \theta)$. Then there is a natural bijection between irreducible m-representations of $H = A \times' G$ whose spectra are $\subset G\{\chi\}$ (the orbit of χ under the affine action) and irreducible m_0-representations of G_χ, the stabilizer of χ in G for the affine action. If the affine action is regular, every irreducible m-representation of H, up to unitary equivalence, is obtained by this procedure.*

Corollary. *Suppose $H^1(G, A^*) = 0$. Then we can take $\theta(g) = 1$ and $m(ag, a'g') = m_0(g, g')$. In this case, the affine action reduces to the ordinary action.*

Remark. It follows easily from the relationship between the affine actions defined by two equivalent cocycles θ, θ' described earlier that the classes of PURs defined by θ and θ' are equivalent.

6.3 Elementary particles over nonarchimedean fields

From our point of view the simplest situation arises when spacetime is a finite dimensional, isotropic quadratic vector space over a nonarchimedean local

field k of ch $\neq 2$. We shall actually assume that $k = \mathbf{Q}_p$. Let $G = \mathrm{SO}(V)$ be the group of k-points of the corresponding orthogonal group preserving the quadratic form. By the k-Poincaré group we shall mean the group

$$P_k = V \times' G.$$

It is the group of k-points of the corresponding algebraic group which is defined over k. Note that when $k = \mathbf{R}$ this means that we do not get the connected Lorentz group but the group where only the condition that the determinant is 1 is imposed, not the orthochronocity. The requirement of Minkowki signature does not make sense over k. Instead we fix the *Witt class*. The results below do not depend on the Witt class. We write V' for the algebraic dual of V. With its k-topology it becomes isomorphic with V^* as a $\mathrm{GL}(V)$-module.

We have already seen that all PURs of the Poincar'e group have trivial multipliers when restricted to the translation subgroup V. The main point here therefore is that the cohomology $H^1(G, V') = 0$, as was proved in [17]. Hence the PUIR's of P_k can be obtained by the above theorems. They are classified by the orbits of G in V'. The orbits are: the level sets of the quadratic form when the value of the quadratic form (called the *mass*) is nonzero; the level set of zero with the origin deleted; and the singleton consisting of 0. These are referred to as *massive*, *massless*, and *trivial massless* respectively. The orbit action is regular by the theorem of Effros since all orbits are either closed or open in their closure. All the orbits have invariant measures on them.

This classification shows that the elementary particles over $\mathbf{Q}_p$ have a richer structure than in the real case. The PUIR's are first classified by mass, and for a given mass, by the PUIR's of the stabilizer in G of a point in that mass orbit. Unlike the real case we cannot replace the PUIRs of the little groups by URs of a single TCE of these groups. The determination of all the multipliers of the little groups is still not completely done.

Study of $H^1(G, U)$ for p-adic Lie groups G

So everything boils down to the vanishing of $H^1(G, V')$. This is not completely trivial and so let me try to explain it in some detail. For various technical aspects of p-adic Lie groups and 1-cocycles see [22]. We look at $H^1(G, U)$ where G is a p-adic Lie group and U a finite dimensional vec-

tor space over $\mathbf{Q}_p$ on which G acts linearly. The cocycles are morphisms of p-adic Lie groups and so are analytic and one can differentiate. For any $X \in \mathfrak{g} = \mathrm{Lie}(G)$ let $\partial(X)$ be the left invariant vector field on G differentiating by the rule $(\partial(X)g)(x) = (d/dt)_{t=0} g(x \exp tX)$ for $x \in G$ and analytic functions g on open sets of G containing x. Then we have a linear map

$$X \longmapsto (\partial f)(X) := (\partial(X) f)(1).$$

If we replace f by $f + f'$ where $f'(g) = g \cdot u - u$ for some $u \in U$, then $\partial f'(X) = \partial f(X) + Xu$ showing that the cohomology class of ∂f does not change. So $f \mapsto \partial f$ induces a map

$$\partial : H^1(G, U) \longrightarrow H^1(\mathfrak{g}, U).$$

For any cocycle f we write $[f]$ for its cohomology class. Notice that as the group cocycles are U-valued, they become a vector space over $\mathbf{Q}_p$ and so $H^1(G, U)$ is a $\mathbf{Q}_p$-vector space and ∂ is a $\mathbf{Q}_p$-linear map.

In general the map $H^1(G, U) \longrightarrow H^1(\mathfrak{g}, U)$ has no reason to be surjective because a Lie algebra morphism $\mathfrak{g} \longrightarrow U \oplus' \mathfrak{g}$ lifts only to a morphism of an open compact subgroup of $U \times G$, so that the Lie algebra cocycle can only be exponentiated to a cocyle on a compact open subgroup K of G. Moreover one needs the cocycle only on a small neighborhood of 1 in G to get the Lie algebra cocycle. This makes it clear that from the Lie algebra we only get the inductive limit $\lim_K H^1(K, U)$ where K runs through the (diminishing) compact open subgroups of G. For these questions, not only for H^1 but for all H^r see Lazard [23a] and Casselman-Wigner [23b]. We proceed in a simpler way, since we are dealing only with H^1.

The kernel Γ of the map

$$H^1(G, U) \longrightarrow H^1(\mathfrak{g}, U)$$

can easily be computed and it is the set of classes which contain a cocycle vanishing on a compact open subgroup K of G, hence constant on the cosets gK. If f is such a cocycle, it follows that $f(x)$ is fixed by $K \cap xKx^{-1}$. If we now make the technical assumption that the stabilizers of elements of U are eiher of measure 0 in G or all of G, we may conclude that f takes values in U^G, and hence that

$$\Gamma \simeq \mathrm{Hom}_0(G, U^G)$$

where the suffix 0 means morphisms that vanish on some compact open subgroup of G. If this Hom space is 0 and $H^1(\mathfrak{g}, U) = 0$, then $H^1(G, U) = 0$ will follow. The technical assumption, seemingly bizarre, is quite natural. If we assume that the action of G on U is algebraic and defined over k, the stabilizers are algebraic groups defined over k. If G is connected the k-points are then either of measure 0 or the whole of G.

If G is semisimple, the Lie algebra cohomology is 0 as we are in characteristic 0. Hence it is the vanishing of the Hom space that needs to be established. There are several circumstances in which this can be proved. For instance, if G is generated by compact subgroups. Indeed, if K is a compact subgroup and $f \in \mathrm{Hom}_0(G, U^G)$, then $f(K)$ is a finite subgroup of $\mathbf{Q}_p$ and so is 0, so that the kernel of f contains all compact subgroups and so is G. Another condition is if the commutator subgroup of G is of finite index in G; for then $f(G)$ itself is a finite subgroup of $\mathbf{Q}_p$ and so is 0. If G is semisimple this condition is satisfied. for full details see [17].

Galilean group and Galilean particles

Over the reals we have a rather full treatment of the Galilean symmetry and the associated Galilean particles [12, 13]. I shall now discuss how Galilean symmetry can be treated over $\mathbf{Q}_p$. This example is interesting since we have nonzero cohomology. Galilean spacetime $V = k^{r+1}$ has the decomposition into space and time: $V = V_0 \oplus V_1$ where $V_0 = k^r$, $V_1 = k$. The Galilean group is the semi direct product $G = V \times' R$ where R itself is the semi direct product of rotations and boosts. Thus V_0 is a quadratic vector space. We set $R_0 = \mathrm{SO}(V_0)$ at first and set $R = V_0 \times' R_0$. The action of G is defined by

$$r = ((u, \eta), (v, W)) : (x, t) \longmapsto (Wx + tv + u, t + \eta).$$

We write $(\cdot, \cdot)$ for the bilinear form on V_0. The dual V' consists of pairs (ξ, t) with duality $\langle(\xi, t), (u, \eta)\rangle = (\xi, u) + t\eta$. The actions of the group R on V and V' are given by

$$(v, W) : (u, \eta) \longmapsto (Wu + \eta v, \eta), \qquad (v, W) : (\xi, t) \longmapsto (W\xi, t - (W\xi, v)).$$

Let

$$\theta_\tau(v, W) = (2\tau v, -\tau(v, v)) \qquad (\tau \in k, (v, W) \in R).$$

The θ_τ are in $Z^1(R, V')$ and $\tau \longmapsto [\theta_\tau]$ is an isomorphism of k with $H^1(R, V')$. Let n_0 be a multiplier for R_0 and let n be the lift to G of n_0 via the composition of the maps $G \longrightarrow R$ and $R \longrightarrow R_0$. Define $m_{n_0,\tau}$ by

$$m_{n_0,\tau}(r, r') = n_0((v, W), (v', W'))\psi(-2\tau(v, Wu') - \tau\eta'(v, v))$$

for $r = ((u, \eta), (v, W))$, $r' = ((u', \eta'), (v', W')) \in G$. Then it follows from [17] that

$$(n_0, \tau) \longmapsto nm_{n_0,\tau}$$

gives an isomorphism of $H^2(R_0) \times k$ with $H^2(G)$. This extends to the nonarchimedean setting the results of Bargmann [12] and Cassinelli et al [13].

From this we can determine the Galilean particles. The analysis is somewhat involved and we just give the highlights [19]. We treat only the case $\tau \neq 0$. Fix now $\tau \neq 0$. The affine action corresponding to the cocycle θ_τ is given by

$$(v, W) : (\xi, t) \longmapsto (W\xi + 2\tau v, t - (W\xi, v) - \tau(v, v)).$$

It is an easy calculation that the function

$$M : (\xi, t) \longmapsto (\xi, \xi) + 4\tau t$$

is invariant under the affine action and maps onto k, since $M((0, a/4\tau)) = a$. If $M((\xi, t)) = a$ the element $(\xi/2\tau, I)$ of R sends $(0, a/4\tau)$ to (ξ, t). Hence the sets M_a where M takes the value a are orbits for the affine action. The stabilizer in R of $(0, a/4\tau)$ is just R_0. Hence for a given a the $m_{n_0,\tau}$-representations are parametrized by the n_0-representations of R_0.

However a little more analysis reveals that *for different a all these representations are projectively the same.* The projection map

$$(\xi, t) \longmapsto \xi$$

is a bijection of the orbit M_a with V_0'. The action of R on M_a becomes the action

$$\xi \longmapsto W\xi + 2\tau v$$

under this bijection and so Lebesgue measure is invariant. *The parameter a has disappeared in the action.* In the Hilbert space of the corresponding representation, the spacetime translation (u, η) acts as multiplication by

$$\psi\left((u,\xi)+\frac{\eta(a-(\xi,\xi))}{4\tau}\right).$$

The factor

$$\psi\left(\frac{\eta a}{4\tau}\right)$$

pulls out and is independent of the variable ξ. Hence it is a phase factor and can be omitted. *The resulting projective representation is thus independent of* a. Hence all these representations represent a single particle. The true parameters are $\tau(\neq 0)$ and the projective representations μ of R_0. We interpret τ as the *Schrödinger mass*, and μ as the *spin* [19].

Conformal symmetry

It was discovered long ago [24a, 24b] that Maxwell's equations are invariant under conformal transformations of spacetime. In recent years this question has been treated from the point of view of representation theory. The starting point of this view is the fact that the Poincaré group P is imbedded in the conformal group C which is SO(2, 4). If now U is an UIR of P, it makes sense to ask if U *has an extension* to C as a UR and if so, is the extension unique? Such an extension will obviously be a UIR of C. The conformal group includes dilations and so for an extension to exist, the particle represented by U has to be massless. It was shown by many people (see the references in [25]) that the UIRs representing the *massless particles with finite helicity* extend as URs of the conformal group. The question makes sense and is interesting in all dimensions. Over the reals the Poincaré group P_{n+1} of the Minkowski space $\mathbf{R}^{1,n}$ can be imbedded in the *conformal group* $C_{n+1} = \mathrm{SO}(2, n+1)$ in such a way that the spacetime $\mathbf{R}^{1,n}$ is dense and open in a compact homogeneous space for the conformal group, the so-called *conformal spacetime*. In [25] Angelopoulos and Laoues settled this question for all n. The theory of the conformal group goes back to Dirac who was deeply interested in electrodynamics, both classical and quantum, in conformal space; it has been continued in modern times by Angelopoulos, Flato, Fronsdal, Ferrara, Laoues, Sternheimer, and a host of others [25, 26].

It is an interesting question as to how much of these theories can be extended to $k = \mathbf{Q}_p$. We begin with the geometry first. Let V a quadratic vector space over k, $\bar{V} = \bar{k} \otimes V$, and P_k (resp. $P_{\bar{k}}$) the k-Poincarģroup of V (resp. the $\bar{k}$-Poincaré group of $\bar{V}$). Let $V_0 = V \oplus U$ where U is a quadratic vector

space with a basis u, v such that $(u, u) = (v, v) = 0, (u, v) = 1$. We define $\bar{V}_0 = \bar{k} \otimes V_0, \bar{U} = \bar{k} \otimes U$. Then $C = \mathrm{SO}(\bar{V}_0)$ is an algebraic group defined over k. We write $C(k)$ for its group of k-points.

Then the group $P_{\bar{k}}$ is isomorphic, as an algebraic group over k to the stabilizer C_u of u in C. The isomorphism is defined over k and gives an isomorphism of P_k with $C_u(k)$, the stabilizer of u in $C(k)$. Writing V_0 as $ku \oplus kv \oplus V$ the imbedding is given (in block matrix form) by

$$(t, R) \longmapsto \begin{pmatrix} 1 & -\frac{(t,t)}{2} & e(t, R) \\ 0 & 1 & 0 \\ 0 & t & R \end{pmatrix}$$

where $e(t, R) \in \mathrm{Hom}(V, ku)$ is the map $v \longmapsto (t, Rv)u$ and $R \in \mathrm{SO}(V)$.

Let Ω be the cone of null vectors in V_0 and $[\Omega]$ its image in projective space. Let $A_u = \{a \in \Omega | (u, a) \neq 0\}$. Then $a = \alpha u + \beta v + w$, where $w \in V$, and $\beta \neq 0$. Taking $\beta = 1$ does not change the image $[a]$ of a in projective space, and then $\alpha = -(w, w)/2$ so that $[a]$ is given by $[-(w, w)/2 : 1 : w]$. Thus $[a]$ is entirely determined by w. Thus $J : w \mapsto [-(w, w)/2 : 1 : w]$ is a bijection of V with the image $[A_u]$ of A_u in projective space. Then one can show the following. There is a natural conformal structure on $[\Omega]$, and the group $C(k)$ acts transitively on $[\Omega]$. Moreover $[A_u]$ is a Zariski open dense subset of $[\Omega]$ stable under C_u, and the imbedding J intertwines the action of the Poincaré group P_k with that of C_u on $[\Omega]$

For $k = \mathbf{Q}_p$, a local field, $[\Omega]$ is compact and so we have a compactification of space-time into $[\Omega]$. For this reason it is natural to call $[\Omega]$ the *conformal space-time* over k. The conformal group in general will move points of spacetime into the infinite part $[\Omega] \setminus A_{|u|}$. It is only the Poincaré group extended by the dilatations that will leave spacetime invariant.

We shall now say that a $\mathbf{Q}_p$-particle has conformal symmetry if the corresponding PUIR extends to the conformal group. It is not difficult to show that massive particles do not have conformal symmetry. However I do not know if massless particles have conformal symmetry. This question, and more generally, the relation between UIRs of the conformal group and the Poincaré group, as well as possible analogs of the Dirac representations and the AdS geometries to the p-adic case remain to be investigated in a systematic manner.

Nonrelativistic quantum mechanics

I would also like to mention a hybrid theory which is really nonrelativistic quantum mechanics in the usual mode, but the kinematical and dynamical objects have p-adic manifolds as sources. For instance, one can have a conventional model where the dynamical Hamiltonian is an operator on the L^2 of a p-adic manifold. The absence of differential operators is not an issue since we can use Fourier transfom theory to define and work with *pseudo-differential operators*. One interesting question is the construction of *path integral representations* for the propagators. The scalar theory was begun in [27a] and the theory for systems with internal degrees of freedom was constructed in [27b]. Unlike the theory over $\mathbf{R}$ where the free hamiltonian is the Laplacian and the path measure is the Wiener measure on the space of continuous paths, in the p-adic case the measure is on the Skorohod space of right continuous paths with only jump discontinuities; it is not a Gaussian measure but the measure associated to a stochastic process with independent increments.

Let K be a nonarchimedean local field and W a finite dimensional vector space over K. We choose a nontrivial additive character on K and define Fourier transforms on functions on W with complex values. Let $\mathbf{F}$ be the Fourier transform operator, defined on $L^2(W)$. We equip W with an ultra metric norm $|\cdot|$ and define the free hamiltonian as $\Delta_b := \mathbf{F} \circ |\cdot|^b \circ \mathbf{F}^{-1}$ where $|\cdot|^b$ is multiplication by $|\cdot|^b$. The full hamiltonian is then $H = \Delta_b + V$ where V is the multiplication by a potential function, also denoted by V. In the vector case the Hilbert space is $L^2(W)^n$ and V is an $n \times n$ matrix valued function. If we work with $\mathbf{R}$ and take $b = 2$ with V having a euclidean norm, then $\Delta_2 = -\Delta$, is the usual Laplacian. Thus the free hamiltonian in the p-adic case is a generalization of the Laplacian. There is no need to restrict to the case $b = 2$.

It turns out that the function $\exp(-t|\xi|^b)$ $(t > 0)$ is the Fourier transform of a probability density function $f_{t,b}$ on W, and the family $(f_{t,b})_{t>0}$ defines a continuous convolution semigroup of probability densities which goes to the Dirac delta measure at the origin as $t \to 0$. Hence for any starting point $x \in V$ we can associate a W-valued separable stochastic process with independent increments, $X(t)$, with $X(0) = x$, such that for any $t > 0.u \geq 0$, $X(t+u) - X(u)$ has the density $f_{t,b}$. It can be shown by standard arguments that go back to Censov in 1956 that the process $X(t)$ defines a probability measure P_x^t in the space D of right continuous paths on $\{t \geq 0\}$ with values in

W having only jump discontinuities; the space D is the one first introduced by Skorohod, provided with a topology that is a complete separable metric. More generally, for given $x, y \in W$ there is a measure $P^t_{x,y}$ on D corresponding to the process $X(t)$ as above, but now conditioned by the requirement that $X(0) = x$, $X(t) = y$. The conditioned measure $P^t_{x,y}$ is constructed in the same way as P^t, using the conditional densities of the joint probability laws at time points $t_1, t_2, \ldots, t_k$.

In [27a] only the scalar case was treated and the potential V was assumed to be bounded also from above, so that there was no question of the essential self-adjointness of the Hamiltonian $H = \Delta_b + V$. In [27b] the vector case was fully treated, V now being a function with values as self-adjoint $n \times n$ matrices. If V is locally bounded, it can be shown that H is essentially self-adjoint on the space of Schwartz–Bruhat functions; the proof in [27b] is essentially that of Kochubei.

The hamiltonian H is now well defined and ≥ 0. So the operator semigroup e^{-zH} ($\Re(z) \geq 0$) is well defined and one can get the quantum propagator e^{itH} by analytic continuation from the positive semigroup e^{-tH} ($t \geq 0$). The Feynman–Kac path-integral formula describes the operator e^{-tH} as an integral operator (propagator) with kernel $K_t(x, y)$ where

$$K_t(x, y) = \int_D \exp\left(- \int_0^t V(\omega(s))ds \right) dP^y_{x,y}(\omega) \, . \, f_{t,b}(x - y).$$

In the vector case one has to be more careful. The integrand in the above formula has to be modified, with a time-ordered exponential in place of the ordinary exponential. The simplest way to proceed is to define, for any a in the Skorohod space of paths with values in $n \times n$ matrices, the time-ordered exponential $E(a)$ by

$$\frac{d}{dt} E(a(t)) = E(a(t))a(t), \qquad (E(a(0)) = I.$$

Then the propagator for e^{-tH} is an integral operator with kernel $K_t(x, y)$ where

$$K_t(x, y) = \int_D \exp\left(- \int_0^t E(-V \circ \omega(s))ds \right) dP^y_{x,y}(\omega) \, . \, f_{t,b}(x - y).$$

For details and references to the works of Censov and Kochubei mentioned above, see [27a, 27b].

Supersymmetry

Finally I would like to mention that the extension of supersymmetry to the p-adic setting is an interesting open question. Supersymmetry over $\mathbf{R}$ is well understood [28a]. For URs of super Lie groups over $\mathbf{R}$ we now have a definitive theory [28b]. Its extension to the nonarchimedean setting is an open challenge. It appears likely that we may have to study representations in p-adic Banach spaces.

6.4 Nonarchimedean quantum field theory

It is consistent with the philosophy expressed in the preceding remarks to ask if one can construct nonarchimedeal analogues of infinite dimensional quantum systems such as quantum fields or scattering theory. This is a very interesting open area and some fundamental results are just being obtained.

First of all, some remarks on usual quantum field theory are in order before we get to the more exotic nonarchimedean case. Historically, the beginnings of the theory of quantized fields goes back to the work of Dirac on radiation theory and of Heisenberg and Pauli who attempted to do canonical quantization of fields in imitation of the standard version of matrix mechanics. After the second world war, when new experiments opened up the problems of theoretical explanations of these results, physicists developed the theory of quantum electrodynamics (QED). In spite of great mathematical difficulties a consistent approach to QED was developed by the heroic efforts of Feynman, Schwinger, and Tomanaga, based on *renormalization*, a technique whose full structure was revealed by the work of Dyson. From then on renormalization was accepted as an essential component of reality, and eventually the *standard model* was constructed, encompassing all known results of elementary particles. The *quantized field* became the fundamental paradigm of modern physics, with the elementary particles arising as the quanta of excited fields [8]. Many people, including Dirac, however had difficulty in accepting many aspects of this development such as renormalization completely, and began searching for more mathematically acceptable ways of reaching the principal results of this developments. This program has had only a very limited success, but nevertheless it has revealed the depth of the questions, and unearthed mathematical problems of great interest, irrespective of their physical

significance, not to mention the foundations it provided for the spin-statistics connection and the PCT invariance.

One approach that has been pursued to give a better understanding of quantum fields is that of the so-called vacuum expectation values

$$E_N(f_1, \ldots, f_N) = (\Phi(f_1) \ldots \Phi(f_N)\Omega, \Omega)$$

where Ω is the vacuum state, and $\Phi(f_j)$ are the field observables, the f_j being test functions on Minkowski spacetime M (either of compact support or rapidly decreasing). The idea of field observables associated to test functions goes back to Bohr and Rosenfeld who showed by decisive arguments that the quantum field strengths fluctuate too wildly so that only their averages in a small region, determined by test bodies, can be regarded as observable in the quantum regime. In his pioneering studies, Wightman essentially defined quantum field theory as the theory of these expectation values E_N, which are distributions on the N-fold product M^N of the spacetime M. The E_N are essentially determined by a sequence of Lorentz invariant distributions W_N on M^N which are really boundary values of analytic functions defined on a tube domain over M^N. The symmetry properties of the E_N are rather complicated because in general one can interchange f_j and f_k only when their supports are *space-like separated*, i.e., when light signals emanating from their supports do not meet. It was Schwinger who took the great and profound step of going over to *Euclidean spacetime* (sometimes also called *Schwinger spacetime*) $\mathcal{E}$, where time has been made pure imaginary, and looking for the vacuum expectation values on $\mathcal{E}^N$. He thus created *Euclidean quantum field theory*. Schwinger's idea was to construct the Minkowski spacetime expectation values by analytic continuation from the Euclidean values. The Euclidean group (at least its nontranslatory part) is *compact*, thus allowing (in his words) this idea to be viewed as an extension of Weyl's unitarian trick to infinite dimensional quantum systems. The Euclidean expectation values actually turn out to be functions, the so-called *Schwinger functions*, on $\mathcal{E}^N$. Heuristically, the Schwinger functions are the moments of a probability measure defined on the space of distributions (or tempered distributions) on the Euclidean spacetime. Thus the construction of quantum fields has been reduced to the construction of certain probability measures on Euclidean spacetime [29a]. For *free fields* the measure is *Gaussian* and so one wants *non-Gaussian* measures to capture the essence of interacting fields.

In their fundamental paper [29b] A. N. Kochubei and M. R. Sait-Ametov have constructed such non-Gaussian measures on the space $\mathcal{D}(\mathbf{Q}_p^n)$ of distributions on $\mathbf{Q}_p^n$. The measures are absolutely continuous with respect to the free field Gaussian measure and live on the spaces of distributions with a fixed compact support (finite volume case). Eventually one has to eliminate this restriction and determine the so-called *infinite volume limit*. The methods used are variations of the methods used in the corresponding real case. They are quite beautiful but technical.

One has to do these problems in all characteristics and over arbitrary local fields as ground fields, perhaps even on very general manifolds. The corresponding scattering theory has been largely untouched. That there must be a serious scattering theory in the nonarchimedean case is clear from the work of Harish-Chandra on the Plancherel formula for reductive p-adic groups, where the Plancherel measure is constructed by a process extremely reminiscent of scattering matrix theory.

References

[1] E. G. Beltrametti, *Can a finite geometry describe the physical space-time*? Universita degli studi di Perugia, Atti del convegno di geometria combinatoria e sue applicazioni, Perugia 1971, 57–62.

[2] E. G. Beltrametti, *Note on the p-adic generalization of Lorentz transformations*, Discrete mathematics, 1(1971), 239–246.

[3] E. G. Beltrametti and Cassinelli, G.,*Quantum mechanics and p-adic numbers*, 2 (1972), 1–7.

[4] I. V. Volovich, *Number theory as the ultimate theory*, CERN preprint CERN-TH.4791/87, 1987.

[5] I. V. Volovich, *p-adic string*, Class. Quantum Grav. 4(1987) L83–L87.

[6] Yu. I. Manin, *Reflections on arithmetical physics*, in *Conformal invariance and string theory,* Acad. Press, 1989, 293.

[7] B. Dragovich, A. Yu. Khrennikov, S. V. Kozyrev, and I. V. Volovich, *On p-adic mathematical physics*, p-Adic Numbers, Ultrametric Analysis, and Applications, 1(2009), 1–17.

[8] I. Yu. Kobzarev, and Yu. I. Manin, *Elementary Particles:Mathematics, Physics, and Philosophy*. Kluwer, 1989.

[9] V. S. Varadarajan, *Geometry of Quantum Theory*, Second Edition, Springer, 2007.

[10] E. Wigner, *On unitary representations of the inhomogeneous Lorentz group*, Ann. Math., 40(1939), 149–204.

[11] H. Weyl, *Group Theory and Quantum Mechanics*, Dover,

[12] V. Bargmann, *On unitary ray representations of continuous groups*, Ann. Math., 59(1954), 1–46.

[13a] G. Cassinelli, E. De Vito, P. Lahti, and A. Levrero, *Symmetries of the quantum state space and group representations*, Rev. Math. Phys., 10(1998), 893–924.

[13b] G. Cassinelli, E. De Vito, P. Lahti, and A. Levrero, *Galilei invariant wave equations*, Reprts. Math. Phys., 43(1999), 467–498.

[13c] G. Cassinelli, E. De Vito, P. Lahti, and A. Levrero, *The Theory of Symmetry Actions in Quantum Mechanics*. Springer Lecture Notes in Physics 654 (2004).
[14] D. Wigner, *Algebraic cohomology of topological groups*, Trans. Amer. Math. Soc., 178(1973), 83–93.
[15] C. C. Moore, *Group extensions of p-adic and adelic linear groups*, Publ. Math. IHES, 35 (1968), 5–70.
[16] G. Prasad and M. S. Raghunathan, *Topological central extensions of semisimple groups over local fields*, Ann. Math., 119 (1984), 143–201; 203–268.
[17] V. S. Varadarajan, *Multipliers for the symmetry groups of p-adic space-time*, p-Adic Numbers, Ultrametric Analysis, and Applications, 1(2009), 69–78.
[18a] G. W. Mackey, *Unitary representations of group extensions.* I, Acta Math., 99 (1958), 265–311.
[18b] G. W. Mackey, *Les ensembles boréliens et les extensions des groupes*, J. Math. Pures Appl. (9) 36 (1957), 171–178.
[19a] V. S. Varadarajan, and Jukka Virtanen, *Structure, classification, and conformal symmetry, of elementary particles over non-archimedean space-time*, Lett. Math Phys., 89 (2009), 171–182
[19b] V. S. Varadarajan, and Jukka Virtanen, *Structure, classification, and conformal symmetry, of elementary particles over non-archimedean space-time*, p-Adic Numbers, Ultrametric Analysis, and Applications, 2(2010), 71–88.
[20] T. Digernes, and V. S. Varadarajan, *Models for the irreducible representations of a Heisenberg group*, Inf. Dim. Anal. Quantum Prob. Rel. Topics 7 (2004)527–546.
[21] E. G. Effros, *Transformation groups and C^*-algebras*, Ann. Math., 81(1965), 38–55.
[22a] J.-P. Serre, *Sur les groupes de congruence des variétés abéliennes*; *Collected papers*, Vol. II, 230–245.
[22b] J.-P. Serre, *Lie Algebras and Lie Groups*, Springer Lecture Notes in mathematics, Vol. 1500, Springer, 2000.
[23a] M. Lazard, *Groupes analytiques p-adiques*, IHES. Math. No. 28 (1965).
[23b] W. Casselman, and D. Wigner, *Continuous cohomology and a conjecture of Serre's*, Inv. Math. 25 (1974), 199–211.
[24a] E. Cunningham, *The principle of relativity in electrodynamics and an extension thereof*, Proc. Math. Soc. London, 8(1909), 77–98.
[24b] H. Bateman, *The transformation of the electrodynamic equations*, 8(1909), 223–264.
[25a] E. Angelopoulos, M. Flato, C. Fronsdal, and D. Sternheimer, *Massless particles, conformal group, and De Sitter universe*, Phys. Rev., D23 (1981), 1278–1289.
[25b] E. Angelopoulos, and M. Laoues, *Masslessness in n-dimensions*, Rev. Math. Phys., 10(1998), 271–300.
[26a] P. A. M . Dirac,*A remarkable representation of the* $3+2$ *de Sitter group*, J. Math. Phys., 4(1963), 901–909.
[26b] M. Flato, C. Fronsdal, and D. Sternheimer, *Singletons, Physics in AdS Universe and Oscillations of composite neutrinos*, Lett.Math. Phys., 48(1999), 109–119.
[26c] S. Ferrara, and C. Fronsdal, *Conformal Maxwell theory as a singleton field theory on ADS_5, IIB three branes and duality*, hep-th/9712239; *Gauge fields and singletons of AdS_{2p+1}*, hep-th/9806072; *Conformal fields in higher dimensions*, CERN-TH/2000-152.
[27a] V. S. Varadarajan, *Path integrals for a class of P-adic Schrödinger equations*, Lett. Math. Phys., 39(1997), 97–106.
[27b] T. Digernes, V. S. Varadarajan, and D. E. Weisbart, *Schrödinger operators on local fields, self-adjointness, and path integral representations for propagators*, Inf. Dim. Anal., Quantum Probability, and Related Topics, 11(2008), 1-18.
[28a] V. S. Varadarajan, *Supersymmetry for Mathematicians:An Introduction*, AMS-Courant Lecture Notes in Mathematics, No. 11, AMS-Courant Institute, 2004.

[28b] C. Carmeli, G. Cassinelli, A. Toigo, and V. S. Varadarajan, *Unitary representations of super Lie groups and applications to the classification and multiplet structure of super particles*, Comm. Math. Phys., 263 (2006) 217–258.

[29a] D. Kazhdan, *Introduction to QFT, Quantum Fields and Strings: A Course for Mathematicians*, Volume 1, AMS and The Institute for Advanced Study, 1999, 377–418.

[29b] A. N. Kochubei, and M. R. Sait-Ametov, *Interaction measures on the space of distributions over the field of p-adic numbers*, Inf. Dimen. Anal. Quantum Prob. Related Topics, 6(2003), 389-411.

Chapter 7
Differential equations with irregular singularities*

Why should an algebraic geometer be interested in
irregular connections?
Pierre Deligne

The first global studies of differential equations with rational coefficients are those of Riemann on the hypergeometric equations. These are special cases of *Fuchsian* equations, or, equations with *regular singularities*. Their theory is essentially controlled by the monodromy action. The equations with *irregular singularities* tell a completely different story. Here the central fact is that formal solutions do not always converge. Their theory goes back to Fabry in 1885 who discovered the phenomenon of ramification, and the decisive developments came from Hukuhara, Levelt, Turrittin, and others. In more recent times, the ideas of Balser, Deligne, Malgrange, Ramis, Subiya, Babbitt and myself, and a host of others, have created a more modern view of irregular linear differential equations that relates them to themes in commutative algebra, linear algebraic groups, and algebraic geometry.

7.1. Introduction
7.2. Reduction theory
7.3. Formal reduction at an irregular singularity and the theory of linear algebraic groups
7.4. Stokes phenomenon
7.5. The Stokes sheaf and the scheme structure on $H^1(St)$
7.6. Reduction theory for families

* This essay owes a great deal to discussions with J.-P. Ramis and Y. Sibuya, to Pierre Deligne through many conversations and letters, and above all, to my long collaboration with Don Babbitt. I am deeply grateful to these and many others for their patience and generosity.

7.1 Introduction

Linear differential equations of the n^{th} order on a domain Ω in the complex plane have been the objects of intense attention from classical times. The most famous of these is the hypergeometric equation

$$z(z-1)\frac{d^2F}{dz^2} + \big[(a+b+1)z - c\big]\frac{dF}{dz} + abF = 0$$

which was studied by Gauss. It has a solution analytic at $z = 0$ given by

$$F(a,b,c:z) = \sum_{n\geq 0} \frac{(a,n)(b,n)}{(c,n)n!} z^n \quad (|z| < 1).$$

Here a, b, c are complex parameters, $c \notin \{0, -1, -2, \dots\}$, and, in the usual notation,

$$(x,n) = x(x+1)\dots(x+n-1) = \frac{\Gamma(x+n)}{\Gamma(x)},$$

Γ being the classical gamma function. Dividing by $z(z-1)$ to make the leading coefficient 1 we see that this equation has 0 and 1 as singular points. Moreover going over to $z = \infty$ by the substitution $z = t^{-1}$ we see that $z = \infty$ is also a singular point. This is an example of a general linear differential equation on $\mathbf{P}^1 = \mathbf{C} \cup \{\infty\}$ with rational coefficients whose poles are contained in a finite set D containing ∞. On any simply connected subdomain of $\mathbf{P}^1 \setminus D$ one has a basis of n linearly independent solutions. It follows from this that if we start with a solution at a point $x_0 \in \Omega$, it can be analytically continued along any path starting and ending at x_0, and not meeting D, but the final solution may well be different from the initial one, although it depends only on the homotopy class of the path. We thus have an action of the fundamental group of $\mathbf{P}^1 \setminus D$ at x_0 on the space of solutions at x_0. This is the *monodromy representation* of the given differential equation.

In a famous paper [1] written in 1857 Riemann undertook for the first time the study of the *global* nature of the solutions for the hypergeometric equation. He recognized the importance of the monodromy representation for this question and computed it explicitly. The fundamental group of $\mathbf{P}_1 \setminus \{0, 1, \infty\} = \mathbf{C} \setminus \{0, 1\}$ is the free group on two generators λ_0, λ_1, the loops in $\mathbf{C}$ around 0, 1 respectively. Riemann determined the 2×2 matrices representing the loops λ_0, λ_1. He restricted himself only to equations whose solutions around a singular point (taken to be $z = 0$) belong to the module (over

the ring of functions holomorphic at $z = 0$) generated by the multi-valued functions $z^a (a \in \mathbf{C})$. The numbers a are the so-called *local exponents* and there are two of them at each of the singular points $0, 1, \infty$; the restriction of Riemann amounts to saying that at each singular point the local exponents do not differ by an integer. Riemann knew that a second order equation with singularities at s, t, u is completely determined by the knowledge of the local exponents of its solutions and introduced the famous notation for the multi-valued function which has the given local exponents:

$$P \begin{pmatrix} s & t & u & \\ \alpha & \beta & \gamma & z \\ \alpha' & \beta' & \gamma' & \end{pmatrix}.$$

For example

$$F(a, b, c : z) = P \begin{pmatrix} 0 & \infty & 1 & \\ 0 & a & 0 & z \\ 1-c & b & c-a-b & \end{pmatrix}.$$

After Riemann it was Fuchs who took up this theme. He considered the class of differential equations on $\mathbf{P}^1$ which have rational coefficients. A singular point of such an equation is called *regular* if the solutions do not have an essential singularity, or, assuming the singular point to be $z = 0$, the solutions have at most polynomial growth: on any open sector with vertex at $z = 0$, the solutions are $O(|z|^{-N})$ for some integer $N \geq 0$ as $z \to 0$. This formulation is necessary because the solutions are in general multi-valued and become single-valued only on simply connected domains, such as a sector. Fuchs considered the equations all of whose singularities are regular; they are called *Fuchsian* nowadays. He discovered the criterion for a differential equation with rational coefficients to be Fuchsian and developed their theory. Eventually this theory was put in definitive form by Frobenius, which is the way we learn the subject nowadays.

That Fuchsian equations arise from algebraic geometry is a very deep fact, the first inkling of which goes back to Legendre and is encoded in his formula

$$\frac{2}{\pi} \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} = F(1/2, 1/2, 1; k^2).$$

And, for

$$K(k) = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

we have the Legendre equation

$$k(1-k^2)\frac{d^2K}{dk^2} + (1-3k^2)\frac{dK}{dk} - kK = 0.$$

The geometric meanon of this equation is that the periods of elliptic integrals satisfy a suitable equation of the hypergeometric type. It was Fuchs who saw that this was not a special happenstance but a general fact, and succeeded in proving that the period integrals of families of Riemann surfaces satisfy regular singular differential equations, thereby inaugurating one of the most profound aspects of our subject, namely the link between meromorphic differential equations and complex algebraic geometry through period integrals. Classical analysts also considered differential equations with doubly periodic coefficients (Lamé's equation) so that already Riemann surfaces other than $\mathbf{P}^1$ were beginning to enter the picture as the manifolds on which the differential equations were defined. The work of Riemann and his successors such as Fuchs, Schwarz, Klein, Poincaré, Pochhammer and others showed that the theory of fuchsian equations was a beautiful structure [2].

Differential equations of the n^h order are equivalent to first order systems

$$\frac{du}{dz} - Au = 0,$$

satisfied by $n \times 1$-vector functions u, A being an $n \times n$ matrix of functions. If we start with the differential equation

$$d^n v + a_{n-1}d^{n-1}v + \cdots + a_1 dv + a_0 v = 0 \qquad d = \frac{d}{dz},$$

the transformation

$$u = \begin{pmatrix} u \\ du \\ \vdots \\ d^{n-1}v \end{pmatrix}$$

leads us to the above system where

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_0 & -a_1 & \dots & -a_{n-1} & 1 \end{pmatrix}.$$

This equivalence is not canonical; for instance we could have chosen

$$u = \begin{pmatrix} u \\ zdu \\ \vdots \\ z^{n-1}d^{n-1}v \end{pmatrix}.$$

The trivial complex vector bundle on Ω has a connection defined on it with the covariant differentiation

$$\nabla_{d/dz} = \frac{d}{dz} - A.$$

The horizontal sections u of this connection are the solutions of the above differential equation and so provide a geometric meaning to the solutions.

The concept of Fuchsian equations can be formulated in the context of several complex variables also. Hypergeometric equations in several variables were considered by Appell and Picard among others. The general theory of regular singular equations in several variables was first treated by Deligne in an extraordinarily influential monograph[3]. In some sense the theory is algebraic geometric. For aspects of the theory in several variables see [4a, 4b].

Nevertheless equations which are not regular singular also arise naturally. This is typically the case when we consider a *family* of regular singular equations where the singularities coincide in the limit with the local data at the singularities changing suitably, or, as it is usual to say, become *confluent*. The *Bessel* equation, the *Whittaker* equation, the *confluent hypergeometric* equation, the *Airy* equation, and many others are examples of differential equations with rational coefficients but with irregular singularities. For instance, we start with the Riemann P-function

$$P\begin{pmatrix} 0 & \infty & c & \\ 1/2+m & -c & c-k & z \\ 1/2-m & 0 & k & \end{pmatrix}$$

and let $c \to \infty$ to get the equation

$$\frac{d^2u}{dz^2} + \frac{du}{dz} + \left(\frac{k}{z} + \frac{1/4 - m^2}{z^2}\right)u = 0$$

which has 0 and ∞ as singularities, with 0 regular and ∞ irregular. If we set $u = e^{-1/2z}W$, the equation for W becomes the *Whittaker equation*

$$\frac{d^2W}{dz^2} + \left(-\frac{1}{4} + \frac{k}{z} + \frac{1/4 - m^2}{z^2}\right)W = 0.$$

Such equations may also be considered on any compact Riemann surface. The equations with irregular singularities give rise to new phenomena; for example, formal solutions seldom converge and solutions typically do not have polynomial growth at a singular point. The general problem is to build a theory of solutions to such equations. Even though the solutions to such equations are very often highly transcendental, classical analysts discovered many equations that fall under this umbrella which were very important in applications. So analysts always have a soft spot for such functions, and it is natural to ask what sort of geometry is described by the solutions of these equations. This geometry is certainly not algebraic and its nature is still very much a mystery. The rhetorical quote at the top of this essay thus makes sense.

Starting from the 1970s, a number of people have been interested in looking at differential equations with irregular singularities. Many letters have been exchanged among some of them. Some of the letters involving Deligne, Malgrange, and Ramis have been collected in [5] and they give a good idea of the themes involved, the possible questions that may be of interest, and so on. Since variations of the periods of complex manifolds leads only to regular singular equations (Griffiths, Deligne) it is not clear what algebraic geometric significance attaches to irregular singular equations; in [5] this issue is taken up in depth.

In this essay I address some questions regarding irregular singular equations which have a group theoretic origin and which have fascinated Babbitt and myself very much for a long time.

7.2 Reduction theory

Even in the theory of Fuchsian equations a large part of the theory revolves around the *local* study of the equations around the singularities. From now

on I shall be concerned almost exclusively with the local theory at a singular point which can be taken as $z = 0$. The central fact about equations is that at any regular singularity, formal solutions are always convergent. At an irregular singularity this is not the case; typically, formal solutions diverge. The classic example goes back at least to Euler who treated the (inhomogeneous) differential equation

$$z^2 \frac{dg}{dz} + g = z$$

whose formal solution is the nowhere convergent series

$$z - 1!z^2 + 2!z^3 - \dots .$$

Examples of homogeneous equations are the Bessel equations which have ∞ as an irregular singularity, and which have divergent expansions there.

The basic idea is to formulate everything in a group theoretic way. So the idea is to try to transform the given differential equation into a simpler one which can be presumably solved right away. Of course one has to restrict the transformations that one is allowed to use. This idea can be formulated in group theoretic terms. If X is the space of differential equations considered and G is the group of allowable transformations, then G acts on X; the technique is then that of finding in each orbit of G in X an element which is especially simple so that the corresponding differential equation may be solved without difficulty. This is an example of *reduction theory* or finding *canonical forms* for the action of G on X, or simply determining $G\backslash X$ in some explicit fashion.

Let us apply this approach to differential equations of the form

$$\frac{du}{dz} = Au$$

where u is an $n \times 1$ vector and A is an $n \times n$ matrix which is meromorphic at $z = 0$, i.e., has an expansion

$$A = A_r z^r + A_{r+1} z^{r+1} + \dots .$$

If the series has a positive radius of convergence A is actually meromorphic in some disk around $z = 0$, otherwise the series is only formal and we say A is *formally meromorphic*. One can transform the differential equation by writing

$$v = gu$$

where g is an invertible meromorphic matrix; the equation for v then becomes

$$\frac{dv}{dz} = Bu \qquad B = gAg^{-1} + \frac{dg}{dz}g^{-1} =: g[A].$$

Let

$$\mathcal{F} = \mathbf{C}[[z]][z^{-1}], \qquad \mathcal{F}_0 = \mathbf{C}[[z]]$$

and let

$$\mathcal{A} = \mathfrak{gl}(n, \mathcal{F}), \qquad \mathcal{G} = \mathrm{GL}(n, \mathcal{F}), \qquad \mathcal{G}_0 = \mathrm{GL}(n, \mathcal{F}_0).$$

Let us denote the convergent elements of $\mathcal{F}$ by $\mathcal{F}_{\mathrm{cgt}}$. Let $\mathcal{A}$ (resp. $\mathcal{A}_{\mathrm{cgt}}$) be the space of formally meromorphic $n \times n$ matrices and $\mathcal{G}$ (resp. $\mathcal{G}_{\mathrm{cgt}}$) the group of invertible elements of $\mathcal{G}$. The above calculation then suggests that we consider the action of $\mathcal{G}$ on $\mathcal{A}$ given by

$$g, A \longmapsto g[A] = gAg^{-1} + \frac{dg}{dz}g^{-1}.$$

Reduction theory is the description of suitable canonical elements in the orbits of the action of $\mathcal{G}$. The formula for the action is reminiscent of gauge actions and so one may think of this as the gauge-adjoint action. Actually we have already mentioned that the solutions of the differential equations may be thought of as the horizontal sections of a connection whose covariant derivative is $d/dz - A$ so that the above action is just the natural action on the space of connections. Although the point of view of connections on vector bundles and gauge transformations is not essential in the local theory, it becomes indispensable in global discussions. The analytic theory corresponds to the action of $\mathcal{G}_{\mathrm{cgt}}$ on $\mathcal{A}_{\mathrm{cgt}}$.

From this point of view the singularity $z = 0$ is regular if A can be reduced to the form

$$z^{-1}C$$

where C is a constant matrix, i.e., there is $g \in \mathrm{GL}(n, \mathcal{F})$ such that

$$g[A] = z^{-1}C.$$

If A is actually meromorphic (not just formally) then g is necessarily in $\mathrm{GL}(n, \mathcal{F}_{\mathrm{cgt}})$. A fundamental (multi-valued) solution to the equation

$$\frac{du}{dz} = Au$$

is

$$F = g(z)e^{(\log z)C}$$

from which we see that the monodromy is

$$e^{2\pi iC}.$$

The solutions are thus in the classical form of meromorphic series multiplied by $z^{\lambda}(\log z)^{r}$. The $\mathrm{GL}(n, \mathbf{C})$-equivalence class of $e^{2\pi iC}$ then determines completely the $\mathcal{G}_{\mathrm{cgt}}$-equivalence class of A. An important *sufficient* condition for regularity is that A has at most a *simple pole* at $z = 0$.

7.3 Formal reduction at an irregular singularity and the theory of linear algebraic groups

The basic system of equations is

$$\frac{du}{dz} = Au. \tag{1}$$

If the matrix A does not have a simple pole at $z = 0$ it is no longer necessarily true that the system of differential equations has a regular singularity at $z = 0$ and so reduction to the form $z^{-1}C$ can no longer be expected. The local reduction theory at an irregular singularity is essentially the creation of many people, notably Hukuhara [6], Katz [7], Levelt [8], Turritin [9]. The group theoretic approach to reduction theory apparently appeared for the first time in [10a] and led to very interesting links with the theory of affine algebraic groups and their actions. I shall briefly discuss this aspect of the theory now.

We start with a matrix A whose entries are in $\mathcal{F}$, the field of formal Laurent series in z. Write

$$A = A_r z^r + A_{r+1} z^{r+1} + \ldots \qquad (r \leq -2, A_p \in \mathfrak{gl}(n, \mathbf{C}), A_r \neq 0) \tag{2}$$

We assume that $r \leq -2$ since the system is Fuchsian for $r \geq -1$. Let $\mathcal{G}_1$ be the subgroup of $\mathcal{G}_0$ of elements $\equiv 1 \bmod z$, i.e., elements of the form

$$g = 1 + zT_1 + z^2 T_2 + \ldots.$$

If P_{μ} are the spectral projections of A_r one can find $g \in \mathcal{G}_1$ such that $B = g[A]$ commutes with all the P_{μ}. The system of differential equations is thus formally *decoupled* along the spectral subspaces of A_r (note that in this reduction A_r is untouched, namely, $B_r = A_r$. If the matrix A_r has n distinct eigenvalues one can then write down n linearly independent formal solutions. But if the matrix A_r is degenerate this is no longer possible. It was Fabry[11] who discovered in 1885 that one should go over to a *covering plane*, namely to work with the variable $\zeta = z^{1/b}$ for some integer $b \geq 2$ and use (formal) meromorphic changes of the *dependent variable* to generate formal solutions. This means that the reduction group is

$$\mathcal{G}_b = \mathrm{GL}(n, \mathcal{F}_b) \qquad \mathcal{F}_b = \mathbf{C}[[\zeta]][\zeta^{-1}].$$

The system (1) is then said to be *ramified*. The fractional Laurent series in $z^{1/b}$ are the *Puisseaux* expansions and the fields $\mathcal{F}_b$ are Galois extensions of $\mathcal{F}$ obtained by adjoining $z^{1/b}$. Of course we know that the algebraic closure $\overline{\mathcal{F}}$ of $\mathcal{F}$ is the union of these extensions, namely,

$$\overline{\mathcal{F}} = \bigcup_{b \geq 1} \mathcal{F}[z^{1/b}].$$

Fabry's idea can now be formulated as the principle that we should work over the algebraic closure of $\mathcal{F}$ for reduction theory. Clearly this principle is a very natural one from the general point of view.

For example, let

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} z^{-2} + \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} z^{-1} \qquad (a \neq 0).$$

If we consider the transformation

$$g = \begin{pmatrix} z^{-2/3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & z^{2/3} \end{pmatrix},$$

then

$$g[A] = B = \begin{pmatrix} 0 & 0 & a \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} z^{-5/3} + \begin{pmatrix} -1/3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/3 \end{pmatrix} z^{-1}.$$

The leading coefficient matrix of B has 3 distinct eigenvalues, namely the three cube roots of $-a$. This determines at once the formal fundamental matrix as a function of $\zeta = z^{1/3}$.

It is of course not obvious how the transformation g was found. Notice that conjugation by g operates on each coefficient $A_k z^k$ of A in two opposite directions, sending some entries of the matrix to those of a higher degree in z and others to ones with a lower degree in z. For this reason the transformation g was classically known as a *shearing transformation*. In the above example, the process stops after one application of a shearing transformation. In the general situation one may have to repeat this process several times with appropriate shearing transformations at each level. The finiteness of the process as well as the method of discovering the sequence of appropriate shearing transformations at the various levels are then not at all obvious.

The reduction procedure is roughly as follows. Since any A can be written as $A^0 + fI$ where $f \in \mathcal{F}$ and A^0 has trace 0, it easily follows that one can confine oneself to the reduction of A^0. Thus we may assume that A has trace 0. If the leading coefficient of A has at least two distinct eigenvalues one can decouple the system of equations along the spectral subspaces of the leading coefficient matrix and so one can use induction on the dimension n. Thus one comes down to the case when the leading coefficient matrix is *nilpotent*.

Let O be the $GL(n, \mathbf{C})$-orbit of A_r where we ae assuming that A_r is nilpotent. The central step in our method is to construct a gauge transformation $g \in GL(n, \mathcal{F}_b)$ for a suitable integer $b \geq 1$ so that the leading term $A'_{r'}$ of $A' = g[A]$ satisfies one of the following:

(a) $A'_{r'}$ has order ≥ -1 and so is regular
(b) $A'_{r'}$ has two distinct eigenvalues
(c) $A'_{r'}$ is nilpotent but lies in an affine subspace of $\mathfrak{gl}(n, \mathbf{C})$ that contains A_r and is *transversal* to the orbit O. Moreover the orbit of $A'_{r'}$ is distinct from that of A_r and has A_r in its closure.

It can now be proved that the properties described in (c) above imply that for the orbit O' of $A'_{r'}$ under $GL(n, \mathbf{C})$ we have

$$\dim_{\mathbf{C}} O' > \dim_{\mathbf{C}} O.$$

Since there are only finitely many nilpotent orbits this inequality makes it clear that the reduction process terminates in a finite number of steps and that it will ultimately end with A being in canonical form (see below for the

precise definition of a canonical form). The result that the orbital dimension of a nilpotent matrix increases when it is moved transversally is to be expected of course, but its proof is not entirely trivial. It depends on general properties of actions of algebraic groups.

I shall now sketch this argument in more detail. The construction of the shearing transformation g above proceeds by first imbedding A_r in a 3-dimensional simple Lie algebra by the use of the well known classical theorem of Jacobson–Morozov according to which we can find elements H and X in $\mathfrak{gl}(n, \mathbf{C})$ such that

$$[H, X] = 2X, \quad [H, Y] = -2Y, \quad [X, Y] = H, \quad Y = A_r.$$

A simple argument shows that if we select a subspace $\mathfrak{a}$ of $\mathfrak{g} = \mathfrak{gl}(n, \mathbf{C})$ which is complementary to the range $[Y, \mathfrak{g}]$ of $\mathrm{ad}Y$, we can find a gauge transformation

$$g = 1 + zT_1 + z^2T_2 + \dots$$

such that

$$g[A] = A_r z^r + A'_{r+1} z^{r+1} + \dots$$

has the property that

$$A'_{r+1}, A'_{r+2}, \dots \text{ are all } \in \mathfrak{a}.$$

One knows from representation theory of $\mathfrak{sl}(2)$ that $\mathfrak{g}_X$ is a possible choice for $\mathfrak{a}$, where $\mathfrak{g}_X$ is the centralizer of X in $\mathfrak{gl}(n, \mathbf{C})$. So we may assume that A itself has this property, namely,

$$A_{r+1}, A_{r+2}, \dots \text{ are all } \in \mathfrak{g}_X.$$

Now, by representation theory of $\mathfrak{sl}(2)$, $\mathfrak{g}_X$ is stable under $\mathrm{ad}H$ and $\mathrm{ad}(H)$ has only *nonnegative integral* eigenvalues there. So we can find a basis $(Z_i)_{1 \le i \le q}$ for $\mathfrak{g}_X$ such that

$$[H, Z_i] = c_i Z_i, \quad c_i \ge 0, c_i \in \mathbf{Z}.$$

The gauge transformation is then of the form z^{qH} where q is a suitable *rational* number. For

$$A' = g[A] = A'_{r'} z^{r'} + \dots$$

we have

$$A'_{r'} = Y + Z, \qquad Z \in \mathfrak{g}_X, Z \neq 0.$$

From now on the argument is entirely geometric. It is a question of showing that if $U \in Y + \mathfrak{g}_X$ is nilpotent and distinct from Y, then the orbit $O_U = \mathrm{GL}(n, \mathbf{C}) \cdot U$ has dimension strictly greater than the dimension of the orbit O_Y of Y. From the theory of orbits of algebraic groups we know that O_U is Zariski open in its Zariski closure $\overline{O_U}$ and that $\overline{O_U} \setminus O_U$ is a union of finitely many orbits of dimension $< \dim(O_U)$. Hence it is enough to prove that $U \notin O_Y$ and Y lies in the closure of O_U in the *usual* topology. If we write

$$Z = c_1 Z_1 + \cdots + c_q Z_q \neq 0, \qquad U = Y + Z$$

then

$$e^{2t} \exp(tH) U \exp -(tH) = Y + \sum_{1 \leq i \leq q} e^{(2+c_i)t} Z_i \to Y \quad (t \to -\infty)$$

For any nilpotent U' one knows that U' is conjugate to cU' for any constant $c \neq 0$; indeed, by Jacobson–Morozov we can write $U' = Y'$ for a standard $\mathfrak{sl}(2)$ triple $\{H', X', Y'\}$, and then

$$\exp(tH') U' \exp -(tH) = e^{-2t} U'.$$

Going back to U which is nilpotent we see that

$$U_t := e^{2t} \exp(tH) U \exp -(tH) \in O_U, \quad U_t \to Y (t \to -\infty).$$

On the other hand, the affine subspace $Y + \mathfrak{g}_X$ is transversal to the orbit O_Y. One can show from this that $Y + \mathfrak{g}_X$ does not meet the orbit O_Y at points close to Y. Since the points U_t of the orbit O_U come close to Y for t very large and negative, we see that O_U is distinct from O_Y. This completes the argument. It is not difficult to see that z^{qH} is essentially the "shearing transformation"occurring in Turrittin's work.

We define a *canonical form* as a matrix of the form

$$B = D_1 z^{r_1} + \cdots + D_m z^{r_m} + z^{-1} C$$

where $r_1 < \cdots < r_m < -1$ are *rational numbers*, $C, D_1, \ldots, D_m$ are elements of $\mathfrak{gl}(n, \mathbf{C})$, the D_j are nonzero and diagonal if $m \neq 0$, with the convention that $B = z^{-1} C$ if $m = 0$ (the regular singular case), and C commutes with all the D_j. The r_j are called the *canonical levels*, r_1 is the *principal level* or the *Katz invariant* (as it was first used by Katz in his beautiful study of

the arithmetic aspects of connections). The connection obtained by omitting the term $z^{-1}C$ is the *irregular part* of B; the canonical form is said to be *unramified* if all the r_j are integers, otherwise *ramified*. Then the main result is that there is a gauge transformation in $z^{1/b}$ for some integer $b \geq 1$ such that

$$g[A] = B, \qquad B \text{ a canonical form}$$

The canonical form is essentially unique except for the changes

$$D_j \mapsto TD_jT^{-1}, C \mapsto TCT^{-1}$$

One may therefore view the canonical form of A as the *complete set of formal invariants of* A. In particular, the canonical levels are fully gauge invariant.

Remark 5. This method generalizes for families as I shall show later. The rational number q in the gauge transformation z^{qH} can be explicitly determined and depends only on the Taylor coefficients A_{r+m} for $0 \leq m < n(|r| - 1)$. The same is true of the irregular part of the canonical form of A. This question was open classically and was settled only after this new method was developed. The bound $M = n(|r| - 1)$ needed for this is sharp.

Remark 6. The group theoretic method described above suggests that the reduction problem (and others also) may be studied in the more general context where $GL(n)$ is replaced by *any complex algebraic group* G and A belongs to the Lie algebra $\mathfrak{g}$ of G; the groups of gauge transformations are the groups of $\mathcal{F}_h$-points of G, namely $G(\mathcal{F}_h)$. The action (2a) of $G(\mathcal{F}_h)$ on $\mathfrak{gl}(\mathcal{F}_h)$ is still well-defined. The canonical forms are defined as before but with the proviso that the B_j, C lie in $\mathfrak{g}$. The method of reduction described above, based on the theory of algebraic groups, goes through with very little change and leads to the same results on reduction and estimates for the number of coefficients needed to determine the canonical form (or at least its irregular part). For all this see [10a, 10b].

Remark 7. I am indebted to Deligne for the remark that the theory over $\mathrm{GL}(n)$ and the fact that the connections form a tensor category are enough to deduce the results of reduction over any algebraic group. This is the so-called *Tannakian* approach to these problems.

7.4 Analytic reduction theory: Stokes phenomenon

I have already remarked on the fact that at an irregular singularity, formal solutions do not always converge, and it is not clear what meaning can be attached to the formal solutions. In particular, the entire reduction theory outlined above hangs in the air without any geometric or analytic significance. The meaning of the formal solutions was discovered by Poincaré in 1896 when he proved that the formal solutions are actually *asymptotic expansions* of analytic solutions at the singularity. Originally such an asymptotic result was proved for solutions of linear differential equations, but eventually it was proved in a very general nonlinear context [12a]. For a more recent exposition, see [12b]. It follows from this that given a formal solution s of a linear meromorphic differential equation with an irregular singularity at $z = 0$, one can find, on any sector with vertex at $z = 0$ and angle sufficiently small, an analytic solution which is asymptotic to s on that sector. If we have a meromorphic connection matrix A with a formal canonical form B (which we assume to be unramified for simplicity, i.e., have integer canonical levels) such that for some gauge transformation $g \in \mathrm{GL}(n, \mathcal{F})$ we have $g[A] = B$, then on sectors with small angles we can find an *analytic* gauge transformation u such that u is asymptotic to g and $u[A] = B$. It should be mentioned that to reach this result the nonlinear form of the asymptotic theorem is essential.

Consider now a connection matrix formally equivalent to a canonical form B. The formal reduction is effected by a gauge transformation g which is typically divergent. In the analytic theory one constructs analytic gauge transformations that are *asymptotic to the formal gauge transformation* g. As we saw above, this can be done on sectors with $z = 0$ as a vertex and with a sufficiently small angle. But, *as one enlarges this sector, the analytic gauge transformation asymptotic to a given formal one will cease to have this asymptotic property when the boundary of the expanding sector crosses certain lines, the so-called Stokes lines.* This is the *Stokes phenomenon*. As the sector moves but with the analytic solution remaining asymptotic to the same formal solution, the fundamental solution gets multiplied by a constant invertible matrices. These constant matrices are the *Stokes multipliers*. The existence of the Stokes phenomenon is clear to see; if the asymptotic property remains the same when the sector expands, the gauge transformation will become meromorphic.

The Stokes phenomenon is a striking example of a discontinuous behavior in a situation where one does not expect such discontinuity. There are many other situations, in the theory of partial differential equations for instance, where such discontinuous asymptotic phenomena are encountered. See the paper of Friedrichs [13] for a very illuminating discussion of such discontinuous asymptotic phenomena, and [12a] for a treatment of many examples of these in the theory of ordinary differential equations. *Unlike the situation with the regular singular equations where monodromy essentially determines everything, it is the Stokes phenomenon that proves decisive for irregular singular equations.*

To make the Stokes phenomenon more precise let us consider a meromorphic connection matrix A and a canonical form B such that for a formal gauge transformation g we have $g[A] = B$. This equation can be rewritten in linear form as

$$\frac{dg}{dz} = Bg - gA$$

so that the asymptotic theorem can be applied. By going over to a suitable covering plane we may assume that everything is over $\mathcal{G} = \mathrm{GL}(n, \mathcal{F})$. We now introduce a covering of the unit circle by sectors Γ_j on each of which there is an analytic gauge transformation g_j asymptotic to the formal gauge transformation g on the sector Γ_j, and $g_j[A] = B$ on Γ_j. If

$$g_{jk} = g_j g_k^{-1}$$

then g_{jk} is asymptotic to 1 on $\Gamma_j \cap \Gamma_k = \Gamma_{jk}$ and

$$g_{jk}[B] = B$$

The g_{jk} are thus *gauge automorphisms* of B that are *asymptotic* to 1 on the overlaps $\Gamma_j \cap \Gamma_k = \Gamma_{jk}$. It is easy to verify that the $\{g_{jk}\}$ verify a *cocycle condition*. The condition that there is a *meromorphic* h such that $h[A] = B$ is that the cocycle be a coboundary. The cohomology class defined by (g_{jk}) is thus an obstruction for the meromorphic equivalence of A and B. This class is an analytic invariant of the given connection A under the group $\mathcal{G}_{\mathrm{cgt}}$.

This cohomological point of view was pioneered by Malgrange and Sibuya who proved [14a, 14b, 15b] the remarkable result that the *Stokes class of a meromorphic connection is a complete invariant of the connection for the group* $\mathcal{G}_{\mathrm{cgt}}$, namely, that $H^1(St(B))$ classifies the systems that are formally

equivalent to B with respect to the relation of *analytic meromorphic equivalence*.

7.5 The Stokes sheaf and the scheme structure on $H^1(St)$

Let us begin with a very classical irregular singular equation, the *Bessel equation*. It is

$$\frac{d^2y}{dt^2} + \frac{1}{t}\frac{dy}{dt} + \left(1 - \frac{v^2}{t^2}\right)y = 0,$$

where $v \in \mathbf{C}$ is a complex parameter. The equation is considered on $\mathbf{P}^1$ and it has $0, \infty$ as singular points, with 0 regular and ∞ irregular. We go to the associated first order system and to $z = t^{-1}$ so that $z = 0$ is the irregular singularity. The system is

$$\frac{du}{dz} = A_v u, \qquad A_v = z^{-2}\begin{pmatrix} 0 & -1 \\ 1 - v^2z^2 & z \end{pmatrix}.$$

We can find $g_v \in \mathrm{GL}(2, \mathbf{C}[[z]])$ such that

$$g_v(0) = \begin{pmatrix} -1 & -i \\ -i & -1 \end{pmatrix}, \qquad g_v[A_v] = B$$

where

$$B = z^{-2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} + z^{-1}\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

The connection matrix B is a canonical form and *does not depend on* v. Thus the Bessel family is *isoformal*. The calculation for g_v shows that $g_v \in \mathrm{GL}(2, \mathbf{C}[v][[z]])$. Thus the reduction to the canonical form depends *holomorphically*, even *polynomially*, on v.

To see what the Stokes sheaf $\mathrm{St} = \mathrm{St}(B)$ is, note that the equation $g[B] = B$ reduces to $g[B_0] = B_0$ where $B_0 = z^{-2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$. The fundamental solution for B_0 is obtained by a trivial integration, namely,

$$\psi = \begin{pmatrix} e^{-i/z} & 0 \\ 0 & e^{i/z} \end{pmatrix}$$

so that $g[B_0] = B_0$ is equivalent to $g\psi = \psi C$ where C is a constant matrix. Thus the Stokes conditions become

$$\psi C \psi^{-1} \sim 1, \qquad C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

or

$$\begin{pmatrix} a & be^{-2i/z} \\ ce^{2i/z} & d \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For z in the upper arc of the unit circle this means that $c = 0$ while for the lower arc we must have $b = 0$. Thus the Stokes sheaf is a sheaf of unipotent groups; the stalk is the upper triangular group T^+ for z in the upper arc, and the lower triangular group T^- for z in the lower arc. At $z = \pm 1$ the stalk is the identity. The Stokes lines are the rays through $z = \pm 1$. We can take the covering consisting of the arcs U^+ where $z \neq -1$, and U^- where $z \neq 1$. It can be shown that this covering is good, i.e., the entire cohomology can be recovered from it. Hence

$$H^1(\mathrm{St}) = T^+ \times T^- \simeq \mathbf{C}^2.$$

On $T^\pm$ we use the nonvanishing off-diagonal entry as the coordinate. A calculation shows that the cohomology class of the Bessel connection A_ν is given by the element of $H^1(\mathrm{St})$ with coordinates

$$(\cos 2\pi\nu, -\cos 2\pi\nu).$$

See [15b].

It is a remarkable fact that for any connection, the corresponding Stokes sheaf is a sheaf of *unipotent subgroups* of $\mathrm{GL}(n, \mathbf{C})$. By looking at specific coverings and using the fact that unipotent group varieties are affine spaces, one can show that the cohomology associated to good coverings is an affine space [14c, 14d]. But the problem of showing that this affine structure is independent of the covering involves technical complications since we are dealing with the cohomology of a sheaf of *nonabelian groups*. In [15b] this question is treated by a method originating from a suggestion of Deligne [5, 15a]. Roughly speaking, we think of the sheaf $\mathcal{S}$ of complex unipotent groups as a sheaf of *unipotent group schemes over* $\mathbf{C}$. This means that for any $\mathbf{C}$-algebra R we have a sheaf $\mathcal{S}(R)$ of groups and we can speak of $H^1(\mathcal{S}(R))$. We thus have a functor

$$R \longmapsto H^1(S(R))$$

and what one should show is that this functor is representable by affine space. This is done in [15b]. The existence of a scheme structure on $H^1(\mathrm{St})$ is quite remarkable.

7.6 Reduction theory for families

In classical analysis many of the natural special transcendentals like the Bessel, or Whittaker functions depend on parameters, and it is a remarkable fact that the *formal* equivalence classes of the members of these families *do not change* with the parameter; indeed we have seen this for the Bessel family above. So these families are *isoformal*, and it is a natural question to analyze the corresponding Stokes phenomena. An important question is whether the formal reduction depends analytically on the parameters. If this can be done, it would follow that the map into $H^1(St)$ taking the connections of the family into their Stokes classes is also analytic, thus making $H^1(St)$ a *moduli space* [15b]. The consideration of such questions reveals a very unexpected and impressive landscape and I want to discuss some of the themes that arise.

To treat such parametric families new methods are needed. The connection matrix A now depends holomorphically on a parameter λ and one wants to find a gauge transformation *also depending holomorphically on* λ that effects the reduction to a canonical form, presumably also depending holomorphically on λ. The approach discussed earlier encounters problems because the Jordan form of the nilpotent matrix which is the leading coefficient of A is now typically *discontinuous* in λ. This is impossible to overcome unless a different approach is used. The new approach is to consider reduction theory not over the complex numbers, but over the *ring of holomorphic functions of* λ, and to generalize the earlier method to this context. Readers who are familiar with algebraic geometry will recognize this as the *relative approach* to reduction theory. The issues are: what are the Jordan forms of nilpotent matrices over rings, whether the Jacobson–Morozov is true over rings, and whether one can prove the increase of dimension of nilpotent orbits under transversal deformations. The theory is somewhat subtle and is treated fully in [10c]. Here we sketch only some high points. For a bird's eye view see [10b].

Jordan forms of nilpotent matrices over rings

The first step is to find the Jordan forms of a holomorphic nilpotent matrix under similarity by invertible holomorphic matrices, or more generally, over fairly arbitrary rings. This question has surprising ramifications. For instance, unlike what happens in the case of complex matrices, there may be *infinitely many nilpotent orbits*, and even *moduli*.

Thus, over the ring $\mathcal{O}_\lambda$ of holomorphic germs at $\lambda = 0$, the matrices

$$Y_m = \begin{pmatrix} 0 & 0 \\ \lambda^m & 0 \end{pmatrix} \qquad (m = 0, 1, 2, \dots)$$

belong to distinct orbits under $\mathrm{GL}(2, \mathcal{O}_\lambda)$. Indeed, all the entries of the matrices of the form UY_mV belong to the ideal generated by λ^m and so Y_m cannot be conjugate to Y_n under $\mathrm{GL}(2, \mathcal{O}_\lambda)$ if $n < m$. Notice that, over $\mathbf{C}$ and for fixed m, all members of the above family for $\lambda \neq 0$ have the Jordan form

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

but at $\lambda = 0$ the matrix is just 0, showing the discontinuity of the Jordan form. The Y_m exhaust the Jordan forms. To see this, note that for a nilpotent L we can find a nonzero element u of $\mathcal{O}_\lambda^2$ in the kernel of L, and we may assume that one of its two components is a unit. We can then find v such that $\{v, u\}$ is a basis for $\mathcal{O}_\lambda^2$. In this basis L has the matrix $\begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix}$, and as $L^2 = 0$ we must have $a = 0$. If we write $c = g\lambda^m$ where g is a unit, conjugacy by $\begin{pmatrix} 1 & 0 \\ 0 & g^{-1} \end{pmatrix}$ reduces L to Y_m.

For 3×3 matrices with Jordan form over the *quotient field* as

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

the Jordan forms over $\mathcal{O}_\lambda$ can be shown to be

$$L_{m_1,m_2,h} = \begin{pmatrix} 0 & 0 & 0 \\ \lambda^{m_1} & 0 & 0 \\ h & \lambda^{m_2} & 0 \end{pmatrix}$$

where h is a polynomial such that

$$\text{degree } (h) < \min(m_1, m_2).$$

Let

$$M_{m_1,m_2} = \{L_{m_1,m_2,h} | \text{degree } (h) < \min(m_1, m_2)\}.$$

Then the orbit space for these nilpotents has two *discrete* parameters m_1, m_2 which are integers ≥ 0, and within each stratum defined by fixing the values of these two discrete parameters, it is a continuum, namely

$$M_{m_1,m_2} \simeq \mathbf{C}^{\min(m_1,m_2)}$$

This situation remains true more or less broadly in general. Within a conjugacy class over the quotient field of $\mathcal{O}_\lambda$, the orbits of nilpotents under $GL(n, \mathcal{O}_\lambda)$ can be first stratified by discrete parameters which we shall call *the discrete invariants*; and once these are fixed, the orbit space becomes a *constructible set*. These remarks suggest that there is much interest in discussing the structure of the adjoint orbits over *function fields* for a semisimple Lie algebra.

Jacobson–Morozov

The theorem of Jacobson–Morozov does not survive over rings. Indeed, if there is a triple $\{H, X, Y = Y_m\}$ over $\mathcal{O}_\lambda$, then H must be of the form $\begin{pmatrix} 1 & 0 \\ c & -1 \end{pmatrix}$ but as $H = XY - YX$, its entries must all be divisible by λ^m. Moreover, for a given Y, even H may not exist: $L_{m_1,m_2,h}$ has an associated H if and only if $h = 0$.

To overcome these difficulties we need to work with invariants for a nilpotent matrix which do not depend on Jacobson–Morozov. Let R be a Noetherian domain with quotient field K, and L a nilpotent matrix, viewed as an endomorphism of R^n. Using a triple $\{H, X, Y = L\}$ *over the field* K we define $W_{t,K}$ as the span of the eigenspaces for H of eigenvalues $\leq t$. Then $(W_{t,K})_{t\in\mathbf{Z}}$ is a filtration, finite and increasing for increasing t with the properties:

(i) $YW_{t,K} \subset W_{t-2,K}$
(ii) If $\overline{W}_{t,K} = W_{t,K}/W_{t-1,K}$, then for all $j \geq 0$ the map L^j induces an isomorphism $L_j : \overline{W}_{j,K} \simeq \overline{W}_{-j,K}$.

The filtration $(W_{t.K})$ is uniquely determined by L and is independent of the choice of H, X. It is the "Hodge filtration"; see [16]. We then define

$$W_t = W_{t.K} \cap R^n.$$

But now L_j is no longer an isomorphism and so we can introduce the R-modules

$$M_j(L) = \overline{W}_{-j}/L_j(\overline{W}_j)$$

which are invariants of L under $\mathrm{GL}(n, R)$.

From general principles of commutative algebra the most important case is when R is a discrete valuation ring. Let us now assume from now on that R is a discrete valuation ring (for example $\mathcal{O}_\lambda$). Then the $M_j(L)$ are finite dimensional vector spaces over the residue field of R and we can define

$$d_j(L) = \dim M_j(L) \quad (j \geq 1).$$

The $d_j(L)$ are the *discrete invariants* of L. For $L = Y_m$ for $n = 2$ we have $d_1 = m, d_j = 0 (j > 1)$; for $n = 3, L = L_{m_1,m_2,h}$ we have $d_2 = m_1 + m_2, d_j = 0 (j \neq 2)$. For a given nilpotent there is an integer $N \geq 1$ such that $d_j = 0$ for $j > N$ and one can determine the d_j recursively downwards, starting with d_N. Let us define for any nilpotent L, the vector $\mathbf{d}(L)$ by

$$\mathbf{d} = \mathbf{d}(L) = (d_N(L), d_{N-1}(L), \ldots, d_1(L)).$$

We regard $\mathbf{d}(L)$ as an element of $\mathbf{Z}^N$ the latter being given its natural *lexicographic ordering*. The vector $\mathbf{d}(L)$ is an invariant of the $GL(n, R)$-orbit of L.

Transversal nilpotent deformations

It is now natural to ask what happens when we consider an *analytic family* $L(\varepsilon)$ of nilpotents with $L(0) = L$. For this I need to explain what an admissible nilpotent is and what is meant by a strictly transversal deformation of a nilpotent. We shall assume actually that R is a $\mathbf{C}$-algebra and that the residue field of R is $\mathbf{C}$, as in $R = \mathcal{O}_\lambda$.

Since R is a discrete valuation ring the torsion free nature of the modules W_t/W_s $(s < t)$ implies their freeness and thus the existence of submodules $W^{(t)}$ such that $W_t = W_{t-1} \oplus W^{(t)}$; the element $H \in \mathrm{End}(R^n)$ defined by

requiring that $Hu = tu$ for $u \in W^{(t)}$ is said to be *associated to* L. It is not unique since the $W^{(t)}$ are not unique in general. We say that L is *admissible* if there is an H associated to L such that $[H, L] = -2L$.

A given L need not be admissible; however, if we expand L in terms of the eigenspaces in $\mathrm{End}(R^n)$ for $\mathrm{ad}(H)$, it is clear that $L = L_{-2} + L_{-3} + \ldots$ where L_r is in $\mathrm{End}(R^n)$ with $[H, L_r] = rL_r$, from which it follows that $L' := L_{-2}$ is an admissible nilpotent such that H is associated to L', and what is decisive, L' has the same filtration as L and the same torsion modules M_j. From the above remarks it follows that

$$\mathbf{d}(L) = \mathbf{d}(L').$$

Thus admissibility is not such a strong demand and one can work around nonadmissibility in reduction theory. For instance, for $L = L_{m_1,m_2,h}$, admissibility happens if and only if $h = 0$ and we can take $L' = L_{m_1,m_2,0}$.

Strict transversality is a sharpening of transversality over $\mathbf{C}$. Let S be an $\mathrm{ad}(H)$-invariant $\mathbf{C}$-linear subspace of $\mathrm{End}(R^n)$ which is complementary to the range of $\mathrm{ad}(L)$; such an S exists and let S^+ be the subspace of S spanned by the eigenvectors of $\mathrm{ad}(H)$ for the eigenvalues ≥ -1.

Definition. The family $L(\varepsilon)$ of nilpotents ($L(0) = L$) in $\mathrm{End}(R^n)$ is strictly transversal if $L(\varepsilon) \in L + S^+$ for all ε.

The basic theorem of deformations of nilpotent matrices is now the following.

Semicontinuity theorem. *If L is admissible and is moved strictly transversally within the class of nilpotents over R to a nilpotent $L' \neq L$, then either the orbital dimension of L' over the field K is strictly bigger than that of L, or the nilpotents stay in the same orbit over K but the vector $\mathbf{d}(L')$ is strictly less than the vector $\mathbf{d}(L)$ in the lexicographic ordering of $\mathbf{Z}^N$. We denote this situation by*

$$L' \prec\prec L.$$

In analogy with what happens in the case of $\mathbf{C}$ it would be natural to say that the family is *transversal* if $L(\varepsilon) \in L + S$ for all ε. If $R = \mathbf{C}$ it is automatic that $S = S^+$ but this not true always for general R, so that what we have is a sharpening of the transversality condition. As an example of a deformation which is transversal but not strictly, consider $L(\varepsilon) = L_{m_1,m_2,\varepsilon}$ with $m_1, m_2 \geq 1$. Then $L(\varepsilon) = L(0) + \varepsilon Z$ where

$$Z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad H = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \qquad [H, Z] = -4Z.$$

As a consequence of the semicontinuity theorem we have the almost exact analogue of what is true over **C**.

Theorem. *Let notations and assumptions be as above, If* $Z \neq 0 \in S^+$ *and* $L' = L_0 + Z$ *is nilpotent, then* $L' \prec\prec L_0$.

Reduction theory over general rings

We introduce a general ring R, which for simplicity we assume to be a Noetherian integrally closed domain with quotient field F. We have a connection matrix

$$A = A_r z^r + A_{r+1} z^{r+1} + \dots \qquad (r \leq -2)$$

where the A_j are $n \times n$ matrices over R. The aim is to try to get canonical forms for such matrices under the action of gauge transformations which are Laurent series in $z^{1/b}$ with coefficients in R. Inspired by the above discussion for canonical forms for nilpotent matrices we assume that over K the connection has a canonical form. However the canonical form will be defined over the algebraic closure of K and so will bear little resemblance to a connection over R. We therefore assume that the canonical form is really defined over R. Moreover it is quite possible that the different elements of the joint spectrum of the $(D_1, D_2, \dots, D_m)$ may have *contacts* which is an example of the *turning point phenomenon*. We shall say that the connection is *well-behaved* if this is not so.

Definition. Let R be a Noetherian integrally closed domain of characteristic 0, and K be, field of its quotients. Let A be a connection matrix whose Laurent series has coefficients which are matrices over R. We say that A is *well behaved* if it is equivalent to a canonical form (which is also called well behaved)

$$B = D_1 z^{r_1} + \dots + D_m z^{r_m} + z^{-1} C$$

under the group of gauge transformations which are Laurent series in $z^{1/b}$ over the algebraic closure of the field K where

(1) the D_j are diagonal matrices with entries in R

(2) if $(\lambda_1, \ldots, \lambda_m)$ and $(\lambda'_1, \ldots, \lambda'_m)$ are distinct elements in the joint spectrum of $(D_1, \ldots, D_m)$, and if k is the first of the integers i for which $\lambda_i \neq \lambda'_i$, then $\lambda_k - \lambda'_k$ is a unit of R.

The concept of well behaved connection matrices is somewhat stringent. There are examples of connections that are not well behaved and which cannot be reduced to a canonical form under $GL(n, R_{z,\infty})$ where $R_{z,\infty} = \cup_{b\geq 1} R[[z^{1/b}]][z^{-1/b}]$; see [10c], pp. 85–87. Still it is a useful concept, especially for *isoformal families*, namely families where the canonical form over the complex numbers does not change. This is the situation with most classical families—Bessel, Kummer, Whittaker, etc. In fact, we have the following useful result.

Proposition. *Let R be the ring of analytic functions on some fixed polydisk in $\mathbf{C}^d$, and let $A \in \mathfrak{gl}(n, R[[z]][z^{-1}])$ be such that for all λ in that polydisk the complex connection matrices $A(\lambda)$ are unramified and have the same canonical form. Then A is well behaved.*

For well behaved connections on suitably restricted rings one can establish a reduction to canonical form. I give here one of the many possible results.

Theorem. *Let R be a Henselian discrete valuation ring over $\mathbf{C}$ with residue field $\mathbf{C}$, for instance $\mathcal{O}_\lambda$. Let $A \in \mathfrak{gl}(n, \mathcal{O}_\lambda[[z]][z^{-1}])$ be a well behaved connection. Then A can be reduced to a canonical form.*

This result is essentially the *one parameter* case of parametric reduction theory. For rings of analytic functions and function germs it is possible to extend Theorem 1 rather completely.

For analytic families of connections whose canonical forms do not depend on the parameters (*isoformal* families, we have the following result, stated in a sharper form than the preceding theorems, in terms of reduction over $R[[z]]$ itself.

Theorem. *Let Δ be a polydisk in $\mathbf{C}^d$ centered at the origin and let $R(\Delta)$ be the ring of holomorphic functions on Δ. Let*

$$A \in \mathfrak{gl}(n, R(\Delta)[[z]][z^{-1}])$$

be a connection such that $A(\lambda)$ is equivalent to $A(0)$ under $GL(n, \mathbf{C})$ for all $\lambda \in \Delta$. Then there exists a concentric polydisk $\Delta_1 \subset \Delta$ and an $x \in GL(n, R(\Delta_1)[[z]][z^{-1}])$ such that $x[A] = A(0)$.

The proof of the reduction theorem in the henselian case follows closely the proof in the absolute case and uses the semi-continuity theorem and its corollary. The extension to polydiscs needs additional arguments; see [10b, 10c]. The extension of these results to connections that may exhibit tuning points is a major unsolved question. For turning points, see [17].

Of course this survey just touches on those aspects that I have been able to say something about. For other aspects, see the letters in [5].

References

[1] B. Riemann, *Collected Papers*, Springer-Verlag, 1990, 99–119.

[2] J. Gray, *Linear Differential Equations and Group Theory from Riemann to Poincaré*, Birkhäuser, Boston, 1986.

[3] P. Deligne, *Equations Différentielles à points singuliers régulier*, Springer Lecture Notes in Mathematics, Vol. 163, 1970.

[4a] M. Yoshida, *Fuchsian Differential Equations*, Vieweg, 1987.

[4b] V. S. Varadarajan, *Some remarks on meromorphic differential equations with simple singularities*, Calcutta Math. Soc. Diamond Jubilee Volume (1983), 49–61. *Selected Papers*, 427–440, AMS 1999.

[5] P. Deligne, B. Malgrange, and J.-P. Ramis, *Singularités Irréguliers. Correspondances et Documents*, Soc. Math. de France, 2007.

[6] M. Hukuhara, *Selected Papers*, 1997.

[7] N. Katz, *Nilpotent connections and the monodromy theorem: application of a result of Turrittin*, Publ. Math. IHES. 39(1970), 207–238.

[8] A. H. M. Levelt, *Jordan decomposition for a class of singular differential operators*, Ark. Matematik, 13(1975), 1–27.

[9] H. Turritin, *Convergent solutions of ordinary differential equations in the neighborhood of an irregular singular point*, Acta Math., 93(1955), 27–66.

[10a] D. G. Babbitt, and V. S. Varadarajan, *Formal reduction of meromorphic differential equations: a group theoretic view*, Pacific J. of Math., 108(1983), 1–80.

[10b] V. S. Varadarajan, *Linear meromorphic differential equations: a modern point of view*, Bull. AMS., 33(1996), 1–42.

[10c] D. G. Babbitt, and V. S. Varadarajan, *Deformations of nilpotent matrices over rings and reduction of analytic families of meromorphic differential equations*, Mem. AMS., vol. 55, No. 325, 1985, 1–147.

[11] E. Fabry, *Sur les intégrales des équations différentielles lineaires à coefficients rationels*, These, Paris, 1885.

[12a] W. Wasow, *Asymptotic Expansions for Ordinary Differential Equations*, Dover, 1987.

[12b] D. G. Babbitt, and V. S. Varadarajan, *Some remarks on the asymptotic existence theorem for meromorphic differential equations*, J. Fac. Sci. Uni. Tokyo, Ser. IA., 36 (1989), 247–262.

[13] K. O. Friedrichs, *Asymptotic phenomena in mathematical physics*, Bull. AMS., 61 (1955), 485–504.

[14a] Y. Sibuya, *Stokes phenomena*, Bull. AMS., 83 (1977), 1075–1077.

[14b] B. Malgrange, *Remarques sur les équations différentielles à points singulier irréguliers*, in Springer Lecture Notes in Mathematics, Vol. 712, 1979.

[14c] W. Balser, W. Jurkat, and D. A. Lutz, *Birkhoff invariants and Stokes multipliers for meromorphic linear equations*, J. Math. Anal. Appl. 71 (1979), 48–94; *A general*

theory of invariants for meromorphic differential equations, Funkcialaj Ekvacioj, 22(1979), 257–283.

[14d] W. Balser, *Zum Einzigkeitssatz in der invariantentheorie meromorpher Differentialgleichungen*, Jour. Reine und Angewandte Mathematik, 318(1980), 51–82.

[15a] Deligne letter to VSV in [5].

[15b] D. G. Babbitt, and V. S. Varadarajan, *Local Moduli for Meromorphic Differential Equations*, Astérisque, 169–170, 1989

[16] P. Deligne, *La conjecture de Weil*, Publ. Math. IHES, 52 (1980), 137–252.

[17] W. Wasow, *Linear Turning Point Theory*, Applied Mathematical Sciences, 54, Springer-Verlag, 1985.

Chapter 8
Mackey, Harish-Chandra, and representation theory*

May the good things come to us from everywhere.
Rig Veda

Some personal reminiscences of Mackey and Harish-Chandra together with brief comments on their work.

8.1. George Mackey and his view of representation theory
8.2. Harish-Chandra as I knew him
8.3. Some reflections
8.4. Fourier analysis on semisimple groups

8.1 George Mackey and his view of representation theory

I first met George Mackey in the summer of 1961. He came to the University of Washington, Seattle, as a distinguished lecturer and gave twenty lectures, lecturing every day for four weeks. The first half of the series dealt with representation theory and the second half with quantum theory. He was at the peak of his powers and achievements and the lectures were a revelation to me. His imagination, his ability to understand the core issues, and his free and easy charm and accessibility just blew me away. In a very essential sense I became his student, and the impact of that one month has stayed with me for ever. On the occasion of a posthumous retrospective on his work held by the American Mathematical Society I wrote [1a]:

> ... It was a great opportunity for me to see a great master at work, and that encounter shaped my entire mathematical career. I became a great admirer of his

* Over the years I have spoken on many occasions on the work of George Mackey and Harish-Chandra on representation theory, and what it has meant to me. This essay draws upon my articles as well as of my memories.

> way of thinking that encompassed a broad picture of mathematics and physics, and emphasized concepts above brute calculation and ideas above technique. I was deeply impressed by the curiosity as well as humility with which he viewed the role of mathematics and the mathematician in the understanding, description, and interpretation of the world of phenomena around us.

I had many opportunities for long chats with him during this visit. Our topics of discussion were quantum theory and representation theory, the twin themes of his lectures. In quantum theory I was fascinated by his axiomatic approach which took off from von Neumann's theory of the quantum logic as a projective geometry. When I returned to India in 1962 I gave a series of lectures at the Indian Statistical Institute which were then turned into the two Van Nostrand volumes, eventually republished in a second edition, by Springer, as a single volume. The Van Nostrand version was also a single volume; but, Professor Marshall Stone, who was the editor of the series, thought that it was better to split them into two volumes. I must mention that throughout this encounter Professor Stone behaved with perfect courtesy and consideration, making this experience thoroughly gratifying for me.

As for representation theory, I had read parts of Weyl's *Classical Groups* and so was interested in representations of semisimple Lie groups. Mackey's advice was simple: read Harish-Chandra's papers. He said that Harish-Chandra had a "terrifying technique" of Lie algebras and it was essential, for anyone interested in the theory of semisimple groups and their representations to understand this technique. It was an advice that turned around my entire research career and launched me on my long love affair with semisimple groups.

When Calvin Moore organized a celebration of Mackey's 70th birthday he asked me to say a few words. He mentioned that Mackey had always regarded me as a "half student" because of the encounters in Seattle. I was certainly proud of this. Years later when I became interested in the representation theory of super Lie groups, it was his theory of representations of semi direct products that I extended (with my collaborators in Genova) to the supersymmetric context [2]. This work has now been continued by Hadi Salmasian to super nilpotent groups [3].

The articles and reminiscences in [1a, 1b] paint a very vivid and moving picture of Mackey, his mathematics, his almost childlike simplicity, and his profound vision. There is therefore nothing I can add to this portrait except to say that it was a very uplifting experience for me to have such a great

mathematician take interest in a young and unknown person like me and really make genuine suggestions.

He was unique in many ways. As far as I know he did not work under contracts and did not participate in political activities that are common in the world of science. He had a vision of mathematics and physics that was unique and he stuck to it, long after his work became less fashionable. I have always admired him for this attitude. He came up in an era of giants but was able to hold his own quite well. In his understanding of quantum theory, especially from the group theoretic point of view, he was the true successor to von Neumann and Weyl.

Mackey's program, although he never explicitly described it as such, may be described as follows: to determine the unitary irreducible representations (UIR) of all locally compact second countable ((lcsc) groups, and to decompose important representations into irreducible constituents. The origins of the program certainly go back to the result proved in 1943 by Gel'fand and Raikov to the effect that the UIRs of any lcsc group separate its points. This suggested that it would be worth looking at the unitary representation theory for such groups. For compact and abelian groups a substantial theory was already developed in the years 1920-1940.

Many people started the study of representation theory and harmonic analysis for the lcsc groups. But this class of groups seemed too wide for detailed work and only some general results could be obtained. Mackey wisely concentrated his attention on certain specific aspects. His work on the extension of the Stone-von Neumann theorem to general locally compact abelian groups led him naturally to the concept of a system of imprimitivity and the closely related theory of induced representations. These were originally studied for finite groups by Frobenius who was one of the founders of the representation theory of finite groups. Mackey saw that the notions of Frobenius made sense for the general locally compact groups and developed the appropriate generalizations. This proved to be a very fertile ground and led him to the representations of semi direct products, and the classification of their representations in terms of orbits and representations of the stabilizers of points in the orbits. If $G = A \times' H$ is a semi direct product with A abelian, then the UIRs of G are classified by the orbits O of H on $\widehat{A}$, and for a given O and a point $\chi \in O$, by the UIRs of the stabilizer H_χ of χ in H. This classification has a caveat, namely, that the action of H on $\widehat{A}$ be *regular*. This was a new idea, regularity being a measure theoretic restriction: an action of a lcsc group R on a lcsc space X is regular if the orbit space X/R is countably separated,

or what is the same thing, if there are countably many R-invariant Borel functions on X separating the orbits of R. In most natural applications (such as the Poincaré group so important in relativistic physics) one encounters only regular semi direct products. It was left to Effros, a student of Mackey, to prove the beautiful result that regularity is equivalent to the condition that all orbits be locally closed, i.e., open in their closure. Nonregular actions are generally called *ergodic*, for instance the action of $\mathbf{Z}$ on the circle group T by an irrational rotation. If the semi direct product is ergodic, then there are many more UIRs of G than the ones coming from the orbits and stabilizers. The conventional wisdom is that in this case one cannot classify the UIRs, but it may not be too much to hope that partial classifications may be sought for.

This work of Mackey brought out surprising aspects of representation theory of lcsc groups and their relationship to some very sophisticated aspects of measure theory, such as the standard Borel structures. One example of this was the following. Let G be a lcsc group and $\widehat{G}$ be the set of equivalence classes of UIRs of G. Since G is lcsc, all its UIRs are in separable Hilbert spaces, and one can take concrete models of these Hilbert spaces (one for each dimension $n = \infty, 1, 2, \ldots$, say $\ell^2, \mathbf{C}, \mathbf{C}^2, \ldots$) and think of the *set* of all UIRs, say X, in a concrete manner, with a relation $\simeq$ of unitary equivalence. Clearly we can identify $X/\simeq$ with $\widehat{G}$. X is a standard Borel space but $\widehat{G}$ need not be as the relation $\simeq$ may be ergodic. The group G may be said to have a *smooth dual* if $\widehat{G}$ is standard. The Mackey philosophy was that a classification of UIRs of a lcsc group is possible if and only if the group has a smooth dual in this sense. He conjectured that a group has a smooth dual if and only if all of its factor representations are direct sums of irreducibles. Here a factor representation is one whose commutant is a factor in the sense of Murray and von Neumann, namely, its center consists of only the scalars. Such a group is said to be of *type I*. Glimm proved Mackey's conjecture; later other proofs were obtained by Effros. Whether a group is type I or not is a deep question; Harish-Chandra proved that real semi simple Lie groups are type I and Dixmier extended this to include all real algebraic groups. Bernstein proved that reductive groups over local fields are type I. These results however use ideas and techniques far removed from the Mackey universe. The work of Glimm and Effros was very fundamental in the clarification of these ideas. Mackey also showed that for a type I group a very satisfactory theory of decomposition of unitary representations can be built by expressing any representation as a canonical direct integral over the dual. For multiplic-

ity free representations this gives a bijection with measure classes on the dual of the group. Perhaps this work of Mackey will be the foundation for a theory of unitary representations of groups defined over locally compact rings that will unify the theories currently pursued over local fields and adele rings.

His success with the theory of semi direct products led Mackey to the following very general question: if G is a lcsc group and N a closed normal subgroup, how to build the representation theory of G in terms of those of N and $H = G/N$? It is natural to assume that N is of type I and decompose a UIR of G by restricting it to N via projection valued measures on $\widehat{N}$. There is a natural action of H on $\widehat{N}$ and one can expect a good theory if this action is regular. However Mackey found that the analogue of the semi direct product theory can be carried out but one will have to know the theory of *projective* unitary representations of H. This appeared very discouraging at first but Mackey showed that the descent from G to N and H could be accomplished if one works with projective representations from the very beginning. For instance, for low dimensional *nilpotent* groups G one can get a comlete description of $\widehat{G}$. But this approach was superseded when Kirillov discovered a very simple and direct way to construct UIRs of nilpotent *Lie groups* via orbits in the Lie algebra, thus beginning the rather profound connection between orbits and representations. Kirillov's theory was extended in a deep manner to solvable groups by Moore, Pukanszky, Auslander, and Kostant. However it must be pointed out that Mackey did not restrict himself to Lie groups, and his theory subsumed projective representations. A general theory of nilpotent or solvable groups and their duals which might contain the Lie group theory as a special case may still be possible.

His work on projective representations led him to a beautiful theory of central extensions of lcsc groups. He discovered that one has to go beyond continuous cohomology and work in the *Borel category*. His student Calvin Moore pushed Mackey's theory in many remarkable directions, calculating the central extensions of p-adic and adelic linear groups, thereby revealing their connections with classfield theory and arithmetic [4].

Finally, Mackey was the first person to discover that ergodic actions were worth investigating in a major way. He did not push is ideas too strongly, but another of his students, Robert Zimmer, went very far in linking the type of measure theory and group actions that was typical in the Mackey world to questions of great depth and impact in the theory of discrete subgroups of semisimple groups and the work of Margulis [5]. To me, the ideas that link measure theory, Borel cohomology, ergodicity, and so on, to semisimple

groups are among the most surprising that have come out of the work of Moore, Zimmer, and others.

Mackey's ideas also had substantial impact on semisimple groups. His double coset formula for the intertwining numbers of induced representations of finite groups was the inspiration behind Bruhat's work on the irreducibility of the minimal principal series for semisimple groups. Furthermore, Roger Howe who was a student of Calvin Moore, applied Mackey's ideas in combination with his own in a decisive way to develop a beautiful theory of *small* representations of metaplectic groups [6]. Howe's work was continued by his students Jian-Shu Li and Hadi Salmesian to include all classical groups and most, if not all, semisimple groups over local fields of characteristic 0 [7a, 7b].

As I think over Mackey's life and his achievements I have come to realize that he was in many ways an extremely atypical figure in the American academic milieu: very introspective, not very concerned with self-promotion, content to follow his own vision even when it became unfashionable, and working almost to the end of his life trying to communicate his views on a large part of mathematics to the next generation. He had a large number of students, many of them elite mathematicians, who grew up in his world, made beautiful discoveries of their own, and continued his vision through their students. It was my greatest good fortune that I was able to interact with him at a crucial stage in my career.

8.2 Harish-Chandra as I knew him

I came to the mathematics department of Princeton University in January 1960 as a postdoctoral fellow to spend the rest of the academic year there. Sometime during that stay Professor Herbert Robbins of Columbia University invited me to give a talk in his statistics seminar. Robbins was very kind and wanted to know if there was anyone I wanted to meet specially. I suggested Harish-Chandra and he took me to meet him after my talk.

That meeting was very brief as Harish was leaving for a dinner with the Colloquium speaker. He told me that to work in semisimple groups one must know the three volumes of Chevalley backwards (which is an Indian idiom meaning total mastery), and also that I should know "roots." He must have understood at once that I was a novice and refrained from mentioning his

own work. I had already begun looking through the Chevalley volumes and knew that they treated Lie groups not only as transcendental objects but also as objects in the algebraic category. But I did not know what roots meant, nor was I aware of the Séminaire Sophus Lie and Séminaire Chevalley. I would get to know about them in a couple of years.

Even before I left for the US we had a small group consisting of myself, Ranga Rao, K. R. Parthasarathy, and Varadhan, at the Indian Statistical Institute. Actually Varadhan came almost when I was about to leave for the US. Parthasarathy, Ranga Rao, and Varadhan succeeded in developing the complete theory of infinitely divisible probability measures and the associated limit theorems for arbitrary locally compact abelian groups, generalizing the classic work of Kolmogorov and Gnedenko to this context. Varadhan had then extended this theory in a remarkable manner to Hilbert spaces. When I returned to the Indian Statistical Institute in the Fall of 1962, the only member of that group still there was Varadhan. I have discussed elsewhere [8] about that remarkable year of working with him. We made the almost insane, certainly very courageous, decision to see if we could understand Harish-Chandra's work on infinite dimensional representations of semisimple Lie algebras, when the underlying group was *complex*. For infinite dimensional representations coming from the group G it was very clear from his work that the basic object was the pair $(\mathfrak{g}, \mathfrak{k})$ where $\mathfrak{g} = \text{Lie}(G)$ and $\mathfrak{k} = \text{Lie}(K)$, K being a maximal compact subgroup of G. When G is *complex*, the imbedding of $\mathfrak{k}$ in $\mathfrak{g}$ can be specified algebraically (this is already in his papers, even explicitly), so that the infinitesimal theory becomes entirely algebraic. After complexification, $(\mathfrak{g}, \mathfrak{k})$ becomes isomorphic to $(\mathfrak{g}_C \times \mathfrak{g}_C, \mathfrak{g}_C)$, the second factor being *diagonally imbedded* in the first.

Varadhan left for the Courant Institute in 1963 but by that time Ranga Rao and Parthasarathy had come back and we spent the next two years pursuing the project begun with Varadhan. This period, 1963-65, was one of the most wonderful in my career. Calcutta was a very difficult place to live, with recurring power outages, rationing of essentials like rice and sugar, and other disruptions of daily life. The war in Vietnam was heating up and Calcutta, one of the most politicized centers in India, was in an uproar over this. The only constant was our work. We would work nonstop every day. We would read statements by Robert McNamara that it was only a matter of time before the war would end since half of Hanoi's electrical capacity was destroyed. We would laugh at this since we were always working in Calcutta much below capacity and yet the city and the people carried on heroically. I realized at

that time that the US had no understanding of Asia and would always fail in political ventures there, as later and even contemporary events would prove.

Ranga Rao and Parthasarathy kept our work going with their imagination, power, and dedication, and we obtained many new results. We wrote a letter to Harish outlining what we had done but he had already left on a tour of India under the auspices of the University Grants Commission. To illustrate how little the monumental stature of Harish-Chandra was appreciated in India at that time, his visit made absolutely no waves, while the visit of a certain cosmologist, who was staying as a special guest of Lal Bahadur Shastry, the Prime Minister, was played up in the newspapers. To be fair, things are not very different here in the US even today, as the headlines and coverage in even great newspapers like the New York Times have very little correlation to the scientific importance of the topics discussed.

Harish came to Calcutta in January of 1965 and we arranged a visit to our institute. This visit was memorable for us and I have written about it elsewhere [9]:

> ... My impressions of that encounter are still vivid in my mind. He was, to use a sports metaphor, at the top of his game, and was very impatient with the usual delays and other inconveniences typical in an Indian visit. The visit to Calcutta was a part of an extended itinerary arranged (I think) by the University Grants Commission of India. While waiting for the elevator to go to our offices I told him that our work in the preceding two years had led us to a possibly novel way of constructing (infinite dimensional) representations of complex semisimple Lie algebras and hoped that he would listen to what we had done and give his comments. He immediately responded by saying that the infinitesimal approach did not go far enough. I remember Parthasarathy turning to me and mutter in Tamil that I had better handle this tiger! I must say that I was also disturbed at his abrupt manner and so asked him whether, in view of this, we should even bother to proceed to a discussion! He immediately became very gentle and said that he would very much like to know what we had done. Our meeting was wonderful for us and gave a real glimpse into his view of the whole subject. Later on, during a lunch with Professor C. R. Rao and others, I asked him why the exceptional series was important—a question that was to be with him nearly all his life—and he answered that "Selberg has emphasized to me its importance and Selberg is always right." Neither I nor my friends had ever before met a person of such intensity and we were completely overwhelmed by his passion and power. He expressed his views on mathematics as well as everything else with a sharpness and focus that was, at least in our limited experience, unprecedented.

The discussions with him in Calcutta were unforgettable. We were delighted that we had been able to do something interesting. The encounter generated a deep resolve within me to try to understand the method by which he reached the deepest results in the theory, namely, via the differential equations. It occurs to me that in some sense we were very lucky to stumble onto this area and this complex of questions just ten years after his pioneering

efforts started in the early 1950s. It made possible that we could ask reasonable questions and discover new facts before the literature piled up and made progress very difficult.

The next time I met him was in 1968 when I visited the Institute as an Alfred P. Sloan fellow. By that time I had an understanding of the construction of the discrete series and his lectures at the Institute were about the continuous spectrum and the Plancherel formula. His lectures emphasized what he called the philosophy of cusp forms and his method of relating the asymptotics of the matrix coefficients of the various series of representations with the Plancherel formula. But the lectures had to be cut short because he suffered during the second term the first of his many heart attacks. He had been working incredibly hard without breaks for many years, and this pattern of life had caught up with him in a very unpleasant manner. Unfortunately the treatments were relatively primitive even in the US at that time but he did make some changes in his lifestyle—more vacations for a start and no work during the vacations.

I had many opportunities to talk at leisure with him on his work and came to know him quite well. It was during this period that I once asked him if he would suggest some questions that I could look at. He did not want to do it mainly because he felt that I would be comfortable only with those problems that I myself had thought about. This advice to work on what mattered most to me was something I never forgot. From that time on I always tried to work on things that were important to me, no matter how humble they might have seemed to others. The idea of working with super-Schwartz spaces came to me at that time, although I did not do anything till Peter Trombi came to Los Angeles to lend a helping hand. The collaboration with Peter is one of the happiest in my life. We worked out the L^1-version of the theory of spherical functions [10] and discovered the fact that the matrix elements of the discrete series decayed faster and faster as their parameter became larger and larger [11].

If I am not mistaken, the monumental work of Jacquet and Langlands had just appeared. It was during that visit of mine that he spoke admiringly of this work, commenting on how far-reaching it was as they could obtain the Artin nonabelian L-functions by their methods. He thought, and that was the only advice he ever gave me, that perhaps I should study elliptic curves.

I remember once a conversation in which he was telling me how difficult it was for him to learn new things. He said that every time he wanted to go a little deeper into algebraic geometry he was told that he should learn about

schemes. He said that this is like a patient who has an iron deficiency and is asked to swallow a pound of nails every day. It was clear that all his time was spent on his vision of harmonic analysis that required almost inhuman focus and concentration, excluding everything else.

I did not meet with him very often in the years after 1970. One such occasion was at the Williamstown conference in 1972. He had already extended his philosophy of cusp forms, outlined in his Oregon Colloquium lectures, to the p-adic case, and obtained the connection between asymptotics of matrix elements and the Plancherel formula. He did this, both in the real and p-adic case through his wonderful theory of Eisenstein integrals and what he always called the Maass-Selberg relations. Initially he worked with the supercusp forms but later on was able to include all the discrete series. During the next years he had additional heart attacks and his general well-being was deteriorating. He often spoke to me of his internal conflict, imposed by his health, whether to look for new discoveries or polish what he had already done for publication. He chose the former as attested by the masses of unpublished papers he left behind, most of them written only for himself.

In 1982 I received a call from him asking me if I could go to Toronto to deliver his invited talk as he himself was forbidden to travel. I agreed and went to Princeton, spending a week discussing what I should emphasize in my presentation. He had planned to talk on the spectral theory of Whittaker functions. He had done the work in the 1970s, but when he started to organize his work for the talk, he discovered to his horror that the main technical result on which everything hinged on was incomplete: its proof contained a serious flaw. He told me that it was as if "the main beam in his house had fallen." He worked night and day to repair it. He did succeed but the effort triggered another heart attack and he was advised against going to Toronto [12]. It was during this visit that I began to realize that unless he took total rest things would go bad. He himself was aware of this and told me that he had pushed the subject as far as he could, and that it was up to others to continue. The lecture I gave at the AMS meeting was not a success-how could it be otherwise? I spoke to him after the talk by phone and he consoled me saying that things would have been the same if he himself had given the talk. He felt that he had long ago ceased to have any meaningful interaction with the others in the field, and as a consequence, become too isolated.

In the summer of 1983 I received a call from Lily that they would both very much like me and Veda to spend a couple of weeks in Princeton. We went and it was a wonderful time for all of us. I went for long daily walks

with him and we often had dinner together at their house. He had become very introspective and it was clear that his thoughts were far away. Veda gave him a small cross which she had picked up in the Vatican museum and he was very moved by it. I was distinctly uneasy when we came back to Los Angeles. Nevertheless I was unprepared for Helgason's phone call on October 16, 1983 that he was no more.

8.3 Some reflections

Counting from the time I first met him, it is almost fifty years now. Memories are slowly fading, the colors no longer as vivid as when things first happened. I feel that I should set some of my recollections down before the pitiless flow of time erases everything.

His lectures were very inspiring, but only to those who understood his philosophy. He tried hard to explain but he was so far ahead that he was able to reach out only to a small number of people in the audience. The situation was essentially the same when he gave the AMS Colloquium lectures in Oregon in 1970. Bott was the other lecturer in that meeting and Harish was very cognizant of the difference in their lecturing styles. Bott had a natural way of explaining things very simply and was a great raconteur. Harish, with his ascetic temperament, and fresh from a second heart attack, had no sense of the trivial, and his lectures overwhelmed the general audience. The lectures were published [13] in the Bulletin of the AMS and contained a bare outline of a monumental work parts of which are still buried in his unpublished manuscripts.

The last part of his lectures and the Bulletin article deal with the $\mathfrak{p}$-adic groups and represent his first attempt to extend his theory into this domain. However this is somewhat misleading. Already in 1965, when he met us in Calcutta, he told us that one should not stop with the real groups but go to the $\mathfrak{p}$-adic groups as well, and ultimately, to the adelic groups and try to understand and extend the work of Hecke and Siegel in the context of general semisimple groups. He was mathematically close to Selberg and Langlands at that time. When Langlands came up with his epoch-making ideas that are now broadly described as the Langlands program, he was one of the first to understand and emphasize their scope and importance. But he himself was

more conservative, preferring to rely on his own intuition to probe the world where arithmetic and Fourier analysis interacted at a very deep level.

His asceticism and single-mindedness made it very difficult for most people to get close to him. In a world of high flying entrepreneurs of mathematics who behaved like prima donnas and jet-setting conductors, he was the ultimate loner who trusted only his own results and insights. In his mature years there was only one collaboration, the one with Borel on arithmetic subgroups of algebraic groups which led to their famous paper in that subject. This aspect of his professional life became more pronounced in his later years, and had the effect of isolating him from his contemporaries. Like Siegel before him, he saw his work fade into the background as a huge monolith, waiting for the future generation to understand its beauty and power and ability to lead to further discoveries. One cannot say however that this was unexpected. The works of all the great figures of the past have suffered this fate, and he was no exception. Unlike many he would never trumpet his ideas and never insist that the world take notice of it.

As the years passed by, the younger generation did not go back to his ideas. Perhaps his methods were too difficult for anyone else to use. Although this might have been frustrating to him, his philosophical approach to life and work gave him serenity and security. I remember getting very depressed during the Corvallis conference on automorphic forms in 1977 when no one even mentioned his name. This must have upset some others also because I still remember a lecture in which Langlands chided the attitudes of the younger people by pointing out that they were too young to know that there was a time when there was no discrete series.

It is a fact that he made little effort in the introductions to his papers to take the reader along, and that the main results came only after gigantic collections of lemmas. However, in my opinion, the sequence of eight papers [14] leading to the construction of the discrete series form a self-contained whole of great beauty. He must have felt this deeply because of what he said on the occasion of a ceremony at the Indian embassy in Washinton D. C., honoring Indian savants:

> ... The study of continuous groups was started by Sophus Lie. Among these the ones with finite volume are called compact. Around 1925 Hermann Weyl extended the work of Frobenius and Schur to compact groups and discovered a beautiful formula which is now called the Weyl character formula...I joined the field around 1950 and my early papers were concerned with algebraic methods for dealing with infinite dimensional representations. About the same time Laurent Schwartz in France had invented the notion of distributions. It turned out that, in order to handle certain differential equations on the group, his distributions were

just the right concept. With their help it eventually became possible to prove an analogue of Weyl's character formula for all semisimple groups and extend to them the theory of Fourier transforms.

In spite of all the talk about the orbit picture for representations, it was in these papers that the correspondence between orbits and representations was fully established, in the most important case of semi simple groups. The technical power exhibited in these papers, combining analysis, algebra, and geometry, is truly unbelievable. Other approaches have been developed to reach the discrete series, but none has the directness and simplicity of his.

It is hard to tell from his writings precisely when he started to think in a global way about representations and their arithmetic aspects. I am sure it already started in the early 1950s when his close friend Mautner began working with representations of $SL(2, \mathbf{Q}_p)$. Mautner came often to Princeton to talk with him, and it is almost certain that their discussions touched on the p-adic groups. Parallel to these impulses, his interest in number theory probably goes back to the late 1940s when he had attended lectures of Artin on classfield theory, stimulated later by his contacts with Chevalley. He himself had lectured in Columbia on it and also on Siegel's theory of quadratic forms. Borel told me that sometime in the early 1960s Harish excitedly phoned him from a vacation outing about some striking things he had discovered about representations of p-adic groups. By the time of the Williamstown conference in 1972 he had already extended to the p-adic groups the principle he had discovered for real groups that relates the asymptotics of matrix elements to the Plancherel measure. In these lectures he formulated his attitude to the p-adic theory. He called this approach the *Lefschetz Principle*, which insisted that all the primes be treated on an equal footing. For him this meant that (roughly speaking) whatever is true for real groups should be also true for the p-adic groups. His entire work on the p-adic groups is animated by this view [15a, 15b].

With declining health he became less and less interested in publishing his results and simply used the format of announcements. For a man of his achievements the honors accorded to him were few and far between. This is not surprising since he neither promoted himself nor allowed others to do it for him. Occasionally there were suggestions that his methods were very elementary, perhaps implying that the complexity and length of his proofs arose from this. In my opinion such comments are widely off the mark. He had a vision of the entire theory of harmonic analysis at an extraordinarily early stage and spent the rest of his life bringing down that dream to reality.

He was a man in a great hurry and he got to where he wanted to be by whatever means he knew. It was Andre Weil who said it best when he said that "of all people he knew, there were only two for whom there was no such thing as a technical difficulty, namely, Chevalley and Harish-Chandra."

I remember one occasion in 1968 very vividly. He, who was a creature of regular habits, was suddenly forced to eat lunch at the Institute one day, and a bunch of us accompanied him to the Institute cafeteria which was housed at the top floor of the building then. He shared the table with us and the discussion naturally turned to one of his favorite topics—how an artist gives shape to his dreams. He said that in his experience one has a dream that is perfect and crystal-clear but in the process of bringing it down to earth the dream gets mangled so much that at the end one does not know whether to feel relief or elation. I have pondered about this remark many times and marveled at his imagination, which allowed him to dare to dream those great dreams of his, and his power and persistence that made them a reality no matter how long it took.

More than anyone I have come across, he had profound emotional ties with his mathematical quest, perhaps partially explaining his many heart attacks some of which came when he was at crucial stages in his work. Once in 1968 I mentioned to him that some of the work can be simplified somewhat. He did not say anything and I must say I was puzzled a little as I felt this was uncharacteristic of him. But in his lecture a couple of days later he acknowledged his tardiness but blamed it on the fact that he was heart broken that my method bypassed a very beautiful theory of his on certain principal values on the Lie group and its Lie algebra. He then continued, beaming, that he has found a new use for it!

In all my conversations with him he almost never commented on his feelings about his own work and so, when he did so, they stand out very clearly. One of these occasions was about his work on arithmetic subgroups of semisimple groups. He had wanted to prove that these have finite co-volume but he was stuck on generalizing Siegel's ideas from the context of symplectic groups to the general case, for nearly a month. The main difficulty lay in relating the fundamental domains of the group to those of a smaller subgroup. The key idea came to him while he was out for a walk near Columbia University. When he came home he shared his excitement with Lily. He told her that "he had just done some thing that would even have pleased Poincaré." Lily told me that she remembered at the time being struck by his wording. This and the finiteness theorems of orbits gave him

great satisfaction as achievements in a line of thought that had been touched upon by Gauss, Lagrange, Poincaré and Siegel. Such exultations were very uncharacteristic of him.

The real mystery of his creative process, or that of any great artist for that matter, is of course impossible to track down. Indeed there are some instances, some of which I know from personal experience with him, and others by what he himself told me, which are truly startling. The time I mentioned above was also the period when he had run aground in his quest for the discrete series, and his difficulties led to a very serious nervous breakdown. Treatments involving electrical shocks were considered but were finally not advised for him. He became a patient of Dr. Nathan Kline, a pioneer in the use of chemical medications for depression. Dr. Kline's treatment worked for Harish over time. When he started working again, his mind was crystal clear, and he reached his goal by methods entirely different from those that he had been trying before his crisis. The announcement [16] in the *Bulletin of the AMS* describing his results is an extraordinary one, giving the blueprint for the great cycle of papers mentioned earlier [14]. A similar thing happened to him during 1968 when he had the first of his many heart attacks. He was stuck at a crucial point in his theory of p-adic orbital integrals before the heart attack, but after he came home from the hospital, he went around his difficulties by different methods and reached his goals almost effortlessly. I remember his telling me this when I visited him while he was convalescing at his home after returning from his hospital stay.

In spite of his protestations he had a deep feeling for fundamental physics and would often express it in informal discussions. He used to compare the discrete spectrum of a compact arithmetically defined homogeneous space to the phenomenon of diffraction. His treatment of the continuous spectrum is a variation of scattering theory where past and present are replaced by the various chambers defined by the Weyl group with the intertwining operators playing the role of the S-matrices. The Maass–Selberg relations then express the unitarity of the scattering matrices. Nevertheless, perhaps even because of his early experience in physics, he would always express his conviction that fundamental work in physics required a certain intuition that even the most creative mathematicians did not possess. "You may have a Porsche but unless you are careful you may drive it into a ditch" was the way he put it to me once.

He never considered the problem of determining the unitary duals as central to his vision. However he was puzzled by the existence of the exceptional

series which played no role in harmonic analysis on the group and framed his puzzlement by asking "why did God create the exceptional series?" He had discovered that the representations that occur in the regular representation are precisely the tempered ones and in his later years he had come to the conclusion that the ones occurring in the spectrum of a homogeneous space were precisely the ones that were tempered with respect to that homogeneous space. He did not have time to develop this in sufficient detail.

There were very few instances in his writings that were truly introspective. One was the talk he gave in the ICM at Moscow in 1966 [17]:

> ... Although the case of real groups is beginning to be fairly well understood, our knowledge of the p-adic groups is still very rudimentary. Nevertheless there appears to be a deep-seated analogy between these two cases. In my opinion, one of the major tasks confronting us, is to try to discover and comprehend the reasons for this similarity. Once local Fourier analysis is well understood, one would have to globalize the problem by going over to the adele group. It is this global setting, which seems to provide the right framework for the understanding of the work of Hecke and Siegel, that the deeper connections between Fourier analysis and arithmetic are likely to emerge. ...
>
> The algebra $\mathfrak{Z}$ of bi -invariant differential operators plays a very important role in the real case. However, as we have already observed, there does not seem to exist any p-adic analogue for $\mathfrak{Z}$. Nevertheless it appears likely that all the final results of Fourier analysis continue to hold, after some slight reformulation, for the p-adic groups. The unravelling of this mystery, would, in my opinion, be an important achievement.

On the occasion of the 80th birthday of Dirac he had this to say [18].

> ... I have often pondered over the roles of knowledge or experience on the one hand and imagination or intuition on the other, in the process of discovery. I believe that there is a fundamental conflict between the two, and knowledge, by advocating caution, tends to inhibit the flight of imagination. Therefore, a certain naivete, unburdened by conventional wisdom, can sometimes be a positive asset.

Harish-Chandra understood that semisimple groups were objects of complete perfection in an imperfect world, perhaps like the music of Mozart. One needed methods of great power as well as supreme refinement to penetrate their secrets. These he possessed in great abundance, and they powered his life-long quest, through declining health in his last years. He discovered the fundamental concepts and principles: characters and their regularity, temperedness and the Schwartz space, Maass-Selberg relations, and so on. He formulated the Lefschetz principle and tried to continue the work of Siegel to all semisimple groups. With his power he combined a feeling for what was truly beautiful and an awareness for the infinite delicacy of the fundamental results. Taken as a whole, his work, a colossal structure informed throughout by his vision, is unsurpassed in twentieth century mathematics.

8.4 Fourier analysis on semisimple groups

I would like to complete this essay with a very brief sketch of some of the contours of his work. I cannot think of a better way than to begin with some insightful, even moving, comments on his work from some friends and colleagues who had known him intimately and for a long time [19].

> (*Borel*) . . . I heard several times André Weil say that he knew only two mathematicians for whom technical difficulties simply did not exist, namely, Chevalley and Harish-Chandra. However this was technical power at the service of the highest goals. I would be hard put to give another example of a work by a single author extending over 2000 pages at such a consistently high level, or should I say relentlessly high level, devoted to one main goal, broadly conceived. . .
>
> . . . Speaking of exceptional technical powers at the service of high goals brings to mind C. L. Siegel, whose work Harish-Chandra held in great esteem. Once, in a conversation, A. Weil and I realized that, for a number of years, we had both viewed Harish-Chandra as a kind of spiritual heir to Siegel by his mathematical style, power and concentration on very difficult, basic, questions. . .
>
> . . . He often said that semisimple Lie groups are so perfect that they must have a divine origin. . .
>
> . . . In mathematics, Harish's life was indeed a search for fundamental general theorems, with the belief that they should be beautiful, and combine to harmonious theories. He pursued this quest with awesome single-mindedness, persistency, power, and success.
>
> (*Helgason*) . . . His magnum opus within representation theory, appropriately called "Harmonic Analysis on Semisimple Lie Groups," stretches from 1951 (when he considered the first special case) to 1976 when the general Plancherel formula was completed. The cumulative nature of this work of twenty-five years leads one to characterize it as monumental; yet even this does not do justice to the courage and pioneering efforts, conceptual and technical, which were needed to overcome the formidable obstacles along the way. . . .
>
> (*Langlands*) . . . His greatest achievement, the construction of the discrete series, still lay before him, and I have always reckoned it my great good fortune to have met him before 1964, when he proved their existence, and to have been close to him at that time. The discrete series are the keystone of his own theory of harmonic analysis and, it seems to me, to all later developments in representation theory. . .
>
> . . . It is difficult to communicate the grandeur of Harish-Chandra's achievements. The theory he created still stands—if I may be excused a clumsy simile—like a Gothic cathedral, heavily buttressed below, but, in spite of its great height, light and soaring in its upper reaches, coming as close to heaven as mathematics can. . .
>
> (*Mostow*) . . . Each great work in mathematics as in other forms of creativity, has its own uniqueness. But I believe it is accurate to say that no mathematician of our time has successfully completed so long and so arduous a climb, solo.

His interest in infinite dimensional representations of semisimple Lie algebras was generated by a suggestion from Dirac, his supervisor in Cambridge, that he take a look at the Lorentz group. He proceeded at first by the infinitesimal method, as was natural for a physicist. He must have realized that

his methods had an immense range of applicability. Starting from around 1951 he spent the next thirty years and more in building a theory of harmonic analysis on semisimple groups by methods which *never* used classification and were so beautiful because they were so general.

Infinite dimensional representations of semisimple Lie algebras had already entered the picture in a decisive and dramatic manner in the question of constructing by general methods the semisimple Lie algebra corresponding to a Dynkin diagram and the closely related problem of constructing the irreducible finite dimensional representation with a given dominant integral highest weight. The methods that he developed were the ones that would start the theory of Kac-Moody algebras some years later. Some of his results were also obtained independently by Chevalley, in particular the relations, beyond the commutation rules determined by the Dynkin diagram, nowadays called the Serre relations. This theory should be appropriately called the representation theory of the pair $(\mathfrak{g}, \mathfrak{h})$ where $\mathfrak{g}$ is a complex semisimple Lie algebra and $\mathfrak{h}$ is a Cartan subalgebra. The representations concerned are highest weight representations. In general they are infinite dimensional and not related to any representation of the group, but in one important case, namely that of a hermitian symmetric space, some of the highest weight representations are also group representations. I shall come to this a little later.

The representation theory relevant to semisimple *groups* is that of a pair $(\mathfrak{g}, \mathfrak{k})$ where $\mathfrak{g}$ is a real semimple Lie algebra and $\mathfrak{k}$ is a maximal compact subalgebra. For these Harish-Chandra introduced the category of representations consisting of those for which the restriction to $\mathfrak{k}$ decomposes as the sum of irreducible *finite dimensional* representations of $\mathfrak{k}$. Let us call these $\mathfrak{g}$-modules $\mathfrak{k}$-*finite*. By using the classical theory of invariants he proved the completely unexpected result that in any *finitely generated* module of a semisimple Lie algebra $\mathfrak{g}$ which is $\mathfrak{k}$-finite, the isotypical subspaces are finite modules for the center $\mathfrak{Z}$ of the enveloping algebra of $\mathfrak{g}$. He then created his theory of analytic vectors which allowed him to lift his Lie algebra theory to a theory of representations of a real semisiple Lie group G in a Banach space. This would lead to the finiteness of the K-multiplicities of a UIR of G (with finite center) when restricted to its maximal compact subgroup K, the type I property of G, and the existence of characters for irreducible Hilbert space representations with an infinitesimal character, in particular for UIRs. The existence of the character is a consequence of the following result: there is an integer $N \geq 1$ such that for any irreducible unitary representation π of G and any irreducible representation $\mathfrak{d}$ of K, the multiplicity $[\pi : \mathfrak{d}]$ with which

$\mathfrak{d}$ occurs in π satisfies the estimate

$$[\pi : \mathfrak{d}] \leq N \dim(\mathfrak{d}).$$

It follows from this that for any test function $f \in C_c^\infty(G)$ the operator

$$\pi(f) := \int_G f(x)\pi(x)dx$$

is of trace class and

$$\Theta_\pi(f) := \mathrm{Tr}(\pi(f))$$

is a distribution in the sense of Schwartz, satisfying an estimate

$$|\Theta_\pi(f)| \leq ||\partial(u)f||_1$$

where u is an element of the enveloping algebra and $||\cdot||_1$ is the L^1-norm. The character is a distribution which is invariant under inner automorphisms (hereafter called simply an invariant distribution) which determines the representation up to equivalence. If χ_π is the infinitesimal character of π, the character is an invariant distribution for the center $\mathfrak{Z}$ of the enveloping algebra with eigenvalue χ_π, namely,

$$\partial(z)\Theta_\pi = \chi_\pi(z)\Theta_\pi \qquad (z \in \mathfrak{Z}).$$

The character is in this sense an *invariant eigendistribution*.

It was by the infinitesimal method that he obtained his famous result that any irreducible $\mathfrak{k}$-finite representation of the group G is a subquotient of a principal series representation, as also the result that any irreducible representation of the Lie algebra which is *infinitesimally unitary* is the K-finite part of a UIR of the group. He never came back in his later years to the infinitesimal method; but later, when it was revived by ideas from Bernstein-Gel'fand-Gel'fand, Enright, Varadarajan, Wallach, Zuckerman, and others, he felt very gratified. He told me that he felt like a gardener who planted a seed and was fortunate enough to see it grow into a mighty tree.

From then on his emphasis shifted dramatically, and Fourier analysis came to the foreground, with characters, matrix elements, and their differential equations, beginning to play a central role. This is where he left behind most of his contemporaries. No one understood as well as he did that as the groups are noncompact, one has to use *differential* rather than *integral* operators to

penetrate deeper. He had determined the Plancherel formula for *all complex* semisimple Lie groups exploiting brilliantly the fact that they have only one conjugacy class of Cartan subgroups [20a]. The point is that the singular elements of a Cartan subgroup form a variety of *complex codimension* 1 and hence of *real codimension* 2, and so the orbital integrals, smooth on the regular set, *extend smoothly to the whole Cartan subgroup*. Hence one can start with the *limit formula* whose significance had already been underscored by Gel'fand and Naimark, and take its Fourier transform, which gives the Plancherel formula. On the other hand, if the group is real and we fix a Cartan subgroup A, the orbital integrals are smooth on the regular set but in general have jumps across the singular points. Let

$$I_f(a) = \int_{G/A} f(xax^{-1}) d_{G/A}x \qquad F_f(a) = \Delta_A(a) I_f(a) \qquad (a \in A')$$

where A' is the regular set of A and f is a smooth test function on G. Here Δ_A is the Weyl denominator defined with respect to a positive system of roots. Let $\partial(\varpi_A)$ be the differential operator corresponding to the *same* positive system. Then F_f and all its derivatives have only jump discontinuities across singular points of A. Moreover

$$\nabla_A := \partial(\varpi_A) \circ \Delta_A$$

is independent of the positive system. Remarkably, $\nabla_A(I_f)$ extends continuously to the whole Cartan subgroup and for some constant c independent of f we have

$$cf(1) = \nabla_A(I_f)(1) = (\partial(\varpi_A)F_f)(1).$$

This is the limit formula. However c is 0 unless the Cartan subgroup is *fundamental*, i.e., has a compact part of maximal dimension, in particular if the Cartan subgroup is compact. Because of the jumps of F_f one cannot immediately take the Fourier transform of the limit formula. I shall explain later how he overcame this obstacle by working over the *Schwartz space* of the group.

He worked out the Plancherel formula for $\mathrm{SL}(2, \mathbf{R})$ in a beautiful little note in the Proceedings of the National Academy [20b]. It is clear from his method that character values were going to play a fundamental role in the general theory and that the discrete series was the key to everything [21]. The fact that the various series of representations corresponded to the characters

of the various Cartan subgroups must have made it clear in his own mind that discrete series arose only when there was a compact Cartan subgroup. Although all these connections can only be dimly perceived by a reader, they must have resonated within himself with great clarity as one can judge from the force and single-mindedness with which he pursued these ideas.

The construction of the discrete series is arguably among his greatest achievements. The discrete series form the key to the construction of the irreducible representations of a semisimple group that occur in the regular representation, and hence for the problem of obtaining an explicit description of the Plancherel measure.

For a unimodular lcsc group G and a UIR π of it, we say that π *is in the discrete series* if the matrix elements

$$f_{\phi,\psi} : x \longmapsto (\pi(x)\phi, \psi)$$

are in $L^2(G)$. We also say that π is *square integrable*. It is actually enough if just one $f_{\phi,\psi}$ is square integrable for some pair $\phi \neq 0, \psi \neq 0$. Let ω be the equivalence class of π. One then has the orthogonality relations in $L^2(G)$ given by

$$(f_{\phi,\psi}, f_{\phi',\psi'}) = \frac{1}{d(\omega)}(\phi, \phi')(\psi', \psi)$$

where $d(\omega) > 0$ is a constant depending only on ω and is called the *formal degree* of ω. In particular, if ϕ, ψ are unit vectors,

$$\int_G |(\pi(x)\phi, \psi)|^2 dx = \frac{1}{d(\omega)}.$$

If G is compact and we normalize dx by requiring that $\int_G dx = 1$, then $d(\omega)$ is the actual degree of the representation. In the general case these ω are precisely the ones that are direct summands of the regular representation and $d(\omega)$ is the mass of ω for the Plancherel measure. In all cases known (real and $\mathfrak{p}$-adic semisimple groups) the Haar measure can be normalized so that all the $d(\omega)$ are integers, but I do not know how far this can be pushed for general unimodular lcsc groups.

In the general case perhaps more can be said about how far the discrete series go towards getting the UIRs—a question that is certainly worth exploring—but for semisimple groups the importance of the discrete series arises from the fact that even if G does not have a discrete series, one can construct UIRs by inducing from parabolic subgroups which are semi direct

products of the form $P = MU$ where M is reductive with a discrete series and U is unipotent, using discrete series representation of M. Thus getting the discrete series became the over-arching concern for Harish-Chandra.

Already for groups G for which G/K has a G-invariant complex structure and $\text{rank}(G) = \text{rank}(K)$ (so that a maximal torus of K is a compact Cartan subgroup of G), a case that includes $\text{SL}(2, \mathbf{R})$ where G/K is the Poincaré upper half plane, he had constructed highest weight modules for the Lie algebra which integrate to representations of the discrete series, calculated their characters on the compact Cartan subgroup, and their formal degrees. The formulae for the character and formal degree were exactly analogous to the celebrated Weyl formulae [22]. However this method gives nothing when G/K is not hermitian, and even in the hermitian case, only a part of the discrete series could be so obtained, namely, not all characters of the compact Cartan subgroup would correspond to these highestw eight representations. This example must have convinced him that the road to the discrete series lay through the explicit construction of their characters, thus bringing his method in complete harmony with that of Weyl. But characters of infinite dimensional representations exist initially only as distributions, and it was not immediately clear how to speak of their values. At some point he started to believe that the characters were in fact functions.

Although there is more to a character than the fact that it is an invariant eigendistribution, he had the great insight that so far as its local structure is concerned, only the fact that it is an invariant eigendistribution is critical. It took him nearly a decade to overcome all the technical difficulties and prove that *all invariant eigendistributions are functions.* More precisely, they are analytic functions on the regular set, these functions are locally integrable on the entire group, and the eigendistributions are precisely the distributions defined by these functions. Let Θ be an invariant eigendistribution. Let A be a Cartan subgroup, Θ_A the restriction of the character to the regular set of A, and Δ_A, ϖ_A, ∇_A be defined as before. Then the differential equations satisfied by Θ around the singular points imply among other things that

$$\Phi_A := \Delta_A \Theta_A$$

extends analytically to the open set $A'(R)$ where no *real root* is equal to 1. In particular, if $A = B$ is a *compact* Cartan subgroup, then Φ_B is anaytic on all of B and is a linear combination of the Weyl transforms of a character of B. The analogy with the Weyl character formula is clear.

For any action of G on a space X one can formulate the notion of invariant eigendistributions and it is *not always the case* that they are functions. There are two special actions, the one on the group itself by inner automorphisms, and the corresponding one on the Lie algebra. On the Lie algebra one has the algebra of *constant coefficient* invariant differential operators playing the role of $\mathfrak{Z}$. For both of these actions the regularity theorem is true. But what is decisive and absolutely astounding is that the invariant eigendistributions on the group could be obtained (after adjusting for the jacobian of the exponential map) by exponentiation of the invariant eigendistributions on the Lie algebra. The *tempered* invariant eigendistributions on the Lie algebra can be constructed by Fourier analysis as the Fourier transforms of the invariant measures on the orbits, and their exponentiation gives the discrete series on the group.

In the general picture the various Cartan subgroups of G give rise to the various series of representations parametrized by the dual groups of the Cartan subgroups, and so the discrete series would come from the characters of a *compact* Cartan subgroup. The formal dimension formula obtained in the calculations with the hermitian symmetric case pointed to the fact that the characters of the discrete series would correspond to the *regular* characters of the compact Cartan subgroup, namely those which are in general position with respect to the Weyl group. The condition of temperedness on the Lie algebra would translate into the condition that the character values be bounded globally by $|D_G|^{-1/2}$ where D_G is the discriminant of the group.

The fundamental theorem of the characters of the discrete series can be stated now. For simplicity I assume that G is the real form of a complex simply connected semisimple Lie group G_c with a maximal compact U such that $U \cap G = K$ is a maximal compact of G. It is assumed that G and K have the same rank. Let B be a maximal torus of K which is then also a maximal torus of U. Then the Weyl group W_c of G_c operates on B and so on $\widehat{B}$; a character $\xi \in \widehat{B}$ is called regular if its stabilizer in W_c is trivial. We write W_G for the subgroup of W_c that comes from G. Select a positive system of roots for (G, B) and let $\Delta = \Delta_B$ be the corresponding Weyl denominator. If $\mathfrak{b} = \mathrm{Lie}(B)$, then $\widehat{B} \simeq L$ where L is a lattice in $(-1)^{1/2}\mathfrak{b}^*$; we write $\log \xi$ for the element in L corresponding to the character $\xi \in \widehat{B}$. If $H_\alpha \in \mathfrak{b}_c$ is the element dual to the root α then $(\log \xi)(H_\alpha)$ is real. Let

$$\varpi = \prod_{\alpha>0} H_\alpha, \qquad \varepsilon(\xi) = \operatorname{sign} \varpi(\xi) = \prod_{\alpha>0} (\log \xi)(H_\alpha).$$

Let

$$q = (1/2)\dim(G/K).$$

Finally, let $\mathcal{E}_2(G)$ be the set of equivalence classes of the discrete series representations of G.

To specify an eigendistribution on the group it is necessary to describe it on each one of a complete set of mutually nonconjugate Cartan subgroups. Harish-Chandra discovered a surprising variant of this procedure for describing the eigendistributions which would later prove to be the characters of the discrete series. He specified them on a *compact* Catan subgroup and imposed a bound at infinity on the other Cartan subgroups.

The first part of the construction of the discrete series is then the following theorem which specifies the invariant eigendistributions: for each regular $\xi \in \widehat{B}$ there is a *unique* invariant eigendistribution Θ_ξ on G such that

(i) On $B' := B \cap G'$ we have

$$\Theta_\xi = (-1)^q \varepsilon(\xi) \frac{\sum_{s \in W_G} \det(s) \xi^s}{\Delta_B}$$

(ii) $\sup_{G'} |D_G|^{1/2} |\Theta_\xi| < \infty$.

In (ii) D_G is the discriminant function on G; its restriction to any Cartan subgroup A is $\pm\Delta_A^2$ where Δ_A is a Weyl denominator for A. Thus (ii) is the same as saying that on each Cartan subgroup A we have

$$|\Delta_A(a)\Theta_\xi(a)| \leq C \qquad (a \in A \cap G')$$

for some constant C, G' being the set of regular points of G, namely the points where D_G does not vanish. In practical terms $\Delta_A(a)\Theta_x i(a)$ is a sum of characters of A (locally) and (ii) says that only the characters which are bounded enter. This is the decisive condition that allows a continuation of the character from the compact Cartan subgroup to the other Cartan subgroups. The reader should also note that the sum in (i) is only over W_G; if G itself is compact, then $W_G = W_c$ and (i) is the Weyl character formula.

The second part of the construction is the following theorem: for each regular character $\xi \in \widehat{B}$, there is a class $\omega(\xi) \in \mathcal{E}_2(G)$ such that Θ_ξ is the character of $\omega(\xi)$; the map $\xi \longmapsto \omega(\xi)$ is surjective onto $\mathcal{E}_2(G)$, and $\omega(\xi_1) = \omega(\xi_2)$ if and only if there is an element $s \in W_G$ such that $\xi_2 = \xi_1^s$. For the formal degree we have

$$d(\omega(\xi)) = \varpi(\xi) := \prod_{\alpha>0} (\log \xi)(H_\alpha).$$

The analogy with the Weyl formula cannot be any closer. It is also true, although its significance becomes clear only when we go to the continuous spectrum, that each irreducible of K occurs only in finitely many discrete series representations, and the trivial representation of K never occurs. The Harish-Chandra character and formulae are the most beautiful formulae in infinite dimensional representation theory.

To go from the eigendistributions to the discrete series representations is a long path. The essential point is to show that the Fourier components of the Θ_ξ (with respect to right translations by K) are in $L^2(G)$. These Fourier coefficients are K-finite eigenfunctions, and, because of their K-finiteness, satisfy certain additional elliptic equations, and so are analytic functions on G. To prove that they are square integrable on the group requires entirely new ideas on the asymptotics of eigenfunctions.

In his earlier work on spherical functions Harish-Chandra had discovered the method to determine the asymptotic behavior of the spherical eigenfunctions. He now applied the same principles in this new context. However the problem is complicated by the fact that the differential equations now admit solutions that grow too fast at infinity and have nothing to do with harmonic analysis on the group. For example, the matrix elements of *both* the finite dimensional and discrete series representations satisfy the *same* differential equations! To single out the relevant solutions Harish-Chandra introduced the fundamental inequality, the famous *weak inequality*, that the solutions should satisfy.

The idea is very simple. We imitate the situation in the spherical case but to have full generality we work with *matrix spherical functions* or functions f with values in a *finite dimensional* Hilbert space on which K has a unitary action τ from both left and right, satisfying the sphericality condition

$$f(k_1 x k_2) = k_1 \cdot f(x) \cdot k_2 = \tau(k_1) f(x) \tau(k_2)$$

and the usual differential equations. f is said to be τ-*spherical*. If V is one-dimensional and τ is the trivial representation we get the usual spherical functions. We now note that the jacobian in the polar coordinates corresponding to the decomposition $G = K\mathrm{Cl}(A^+)K$ is of the order of $e^{2\rho}$ where 2ρ is the sum of positive roots. In the actual spherical case the eigenfunctions f are just on the verge of being square integrable, namely that $e^{\rho} f$ *has at most*

polynomial growth. This condition is taken over to the general case:

$$|f(a)| \leq Ce^{-\rho(\log a)}(1 + ||\log a||)^r \qquad (C > 0, r \geq 0).$$

A τ-spherical eigenfunction with this property is said to satisfy the *weak inequality*.

To see how f behaves at infinity on G, let $G = KAN$ be an Iwasawa decomposition and A^+ the positive chamber in A defined by N; $A^+ = \exp \mathfrak{a}^+$ where $\mathfrak{a} = \mathrm{Lie}(A)$ and $\mathfrak{a}^+$ is the positive chamber where the roots belonging to N are > 0. Then it is enough, in view of the Cartan decomposition

$$G = K\mathrm{Cl}(A^+)K,$$

(Cl means closure)to examine what happens at infinity on A^+. Now there are several ways of going to infinity on A^+; indeed, if F is a subset of simple roots, then one can insist that only the values of the roots from F go to infinity, while the remaining simple roots stay bounded. The differential equations satisfied by f then have the remarkable property that when we neglect all terms that vanish at infinity (for this choice of F), *they become the spherical differential equations on the reductive group $M(F)$ defined by the simple roots outside F*. In this way the original eigenfunction f is approximated (with a rapidly decreasing error) very well by linear combinations of eigenfunctions on $M(F)$. Harish-Chandra calls this the *constant term f_F of f of along F*. It is essential for this approximation that the eigenfunction f satisfy the weak inequality. For these the approximation by the f_F gives a satisfactory theory. This method applied to *families* of eigenfunctions coming from the various series of representations is the basic tool in the study of the continuous spectrum.

To return to the the Fourier coefficients of the Θ_ξ, it is now a question of showing that they satisfy the weak inequality. This turns out to be a deep question, but once it is taken care of, the analysis sketched above would apply to them. For f to be in $L^2(G)$ the condition is that $f_F = 0$ for all $F \neq \emptyset$, or, as Harish-Chandra calls it, f is a *cusp form.*. If the eigenvalue of f is defined by a *regular character*, a condition satisfied by the Fourier coefficients of Θ_ξ, one can show that $f_F = 0$ for all $F \neq \emptyset$ so that f is a cusp form. Those who are familiar with some structure theory will recognize that the F define parabolic subgroups (cusps) and that the approximation of f is the constant

term of the expansion of f around that cusp. There is thus a clear analogy with the theory of automorphic forms.

It remains to show that the entire discrete series has been obtained. For this one can proceed as Weyl did in the compact case by reducing the analysis to B by integrating over the orbits. However there are other Cartan subgroups and their presence implies that the normalized orbital integrals are *not* smooth but have jumps. Only when f has the property that its orbital integrals over the noncompact Cartan subgroups are 0 can one come down to B. Such functions are actually cusp forms and so one cannot operate inside $C_c^\infty(G)$ but in the *Schwartz space* $\mathcal{C}(G)$ introduced by Harish-Chandra. Since elements of the Schwartz space are *not* in general in $L^1(G)$, their integrability on the conjugacy classes is no longer a given, but has to be proved. Once this is done, the completeness proof can be carried out.

For the complete story of Fourier analysis on semisimple groups that includes the continuous spectrum, there still remains the principle, which he first uncovered in his work on spherical functions, that the coefficients of the asymptotic expansions of the eigenfunctions of the continuous spectrum determine the Plancherel measure in a rather simple manner. For eigenfunction expansions on a half-line this goes back to Weyl [23]. For the special case of the spherical continuous spectrum see [24], where the principle is described as "reminiscent of a result of H. Weyl on ordinary differential equations." The analogy becomes very close when G/K has rank 1 and is in fact the beginning of the analogy with scattering theory alluded to earlier.

The spherical eigenfunctions are given by the integral

$$\varphi(\lambda, x) = \int_K e^{(\lambda-\rho)(H(xk))} dk$$

where H is the function on G defined by

$$x = k \exp H(x) n \qquad (x \in G, k \in K, H(x) \in \mathfrak{a}, n \in N).$$

So we have a *family* of eigenfunctions parametrized by linear functions λ on $\mathfrak{a} = \mathrm{Lie}(A)$ where $G = KAN$ is an Iwasawa decomposition and 2ρ is the sum of the positive roots (restricted) (I use the notation of [24]). By the method of perturbation analysis of the differential equations at infinity on G discussed briefly earlier one can show that at infinity on A^+ when *all* roots go to infinity, we have

$$\varphi(\lambda, a) \sim \sum_s c(s\lambda) e^{(s\lambda)(\log a)}$$

where the sum is over the Weyl group and λ is pure imaginary (unitarity condition). The function c is meromorphic but never 0 when λ is pure imaginary,

$$|c(s\lambda)| = |c(\lambda)|$$

for all s and all pure imaginary λ, and the Plancherel measure is

$$|c(\lambda)|^{-2} d\lambda.$$

Notice that because of the approximation above, if one form a *wave packet*

$$\varphi(u, x) = \int_{(-1)^{1/2}\mathfrak{a}^*} u(\lambda)\varphi(\lambda, x) d\lambda$$

then we can, at least to a first approximation, approximate the wave packet on A^+ by

$$\sum_s \int c(s\lambda) u(\lambda) e^{(s\lambda)(a)} d\lambda$$

and begin to use classical Fourier analysis on A.

In the general case Harish-Chandra uses vector-valued functions with values in a double module for K and parabolic subgroups other than the minimal one for defining integrals analogous to the above one. The spherical integral is clearly analogous to the Eisenstein series one encounters in the theory of automorphic forms, see [25] where he clearly explains this analogy in the introduction; indeed, this is one of the few places in his entire work where he has made a genuine effort to lay bare the intuition and analogies behind his approach. For this reason he called the integrals generalizing the spherical ones the *Eisenstein integrals*. The asymptotics of the Eisenstein integrals and the precise formulation of their relation to the Plancherel measure, not only in the real case but in the p-adic case as well, are his greatest achievements in the realm of the continuous spectrum. Even a brief discussion of this is beyond the framework of this essay.

I am aware that this sketch hardly does any justice to the majestic development in the cycle of papers mentioned earlier [14]. For details I ask the reader to go to my introduction to his collected papers or to [26a, 26b]. In addition, for a good reference to the evolution of Harish-Chandra's ideas in the nearly thirty years since he died, see the articles in the legacy volume cited in

[9]. For the work of Van den Ban, Delorme, Flensted-Jensen, Schlichtkrull, and others on the harmonic analysis of semisimple symmetric spaces, see [27a, 27b].

References

[1a] R. S. Doran, C. C. Moore, and R. S. Zimmer, (Eds.), *George Mackey and his work on representation theory and foundations of physics. Group representations, ergodic theory, and mathematical physics: a tribute to George W. Mackey*, 417–446, Contemp. Math., 449, Amer. Math. Soc., Providence, RI, 2008.

[1b] R. S. Doran, and A. Ramsay, *George Mackey 1916-2006*, Notices of the Amer. math. Soc., 54(2007), 824-850.

[2] C. Carmeli, G. Cassinelli, A. Toigo, and V. S. Varadarajan, *Unitary representations of super Lie groups and applications to the classification and multiplet structure of super particles*, Comm. Math. Phys., 263(2006), 217-258.

[3] H. Salmasian, *Unitary representations of nilpotent super Lie groups*, Comm. Math. Phys., 297(2010), 189–227.

[4] C. C. Moore, *Group extensions of p-adic and adelic linear groups*, Pub. Math. IHES, 35(1968), 5–70.

[5] R. J. Zimmer, *Ergodic Theory and Semisimple Groups*, Birkhäuser, 1984.

[6] R. Howe, *On a notion of rank for unitary representations of the classical groups*, Harmonic analysis and grouprepresentations, 223-331, Liguori, Naples, 1982.

[7a] J,-S. Li, *On the classification of irreducible low rank unitary representations of classical groups*, Compositio Math. 71 (1989), no. 1, 29-48.

[7b] H. Salamsian, *A notion of rank for unitary representations of reductive groups based on Kirillov's orbit method*. Duke Math. Jour. 136 (2007), 1-49.

[8] V. S. Varadarajan, *Some mathematical reminiscences*, Methods Appl. Anal. 9 (2002) v–xviii.

[9] V. S. Varadarajan, *Harish-Chandra, His Work, and its Legacy*, in *The Mathematical Legacy of Harish-Chandra*, 1–35, R. S. Doran, and V. S. Varadarajan, (Eds.), Amer. Math. Soc., Proc. Symp. in Pure Math., Vol. 68, 2000.

[10] P. C. Trombi, and V. S. Varadarajan, *Spherical transforms of semisimple Lie groups*, Ann. of Math. (2) 94(1971), 246–303.

[11] P. C. Trombi, and V. S. Varadarajan, *Asymptotic behaviour of eigen functions on a semisimple Lie group: the discrete spectrum*, Acta. Math. 129 (1972), 237–280.

[12] J. A. C. Kolk, and V. S. Varadarajan, *On the transverse symbol of vectorial distributions and some applications to harmonic analysis*, Indag. Math. (N.S.) 7 (1996), 67–96. [13] *Harish-Chandra Collected Papers*, Vol. IV., [1970c], Springer -Verlag, 1984.

[14] *Harish-Chandra Collected Papers*, Vol. I., [1964a]–[1966b], Springer -Verlag, 1984.

[15a] *Harish-Chandra Collected Papers*, Vol. IV., [1973], Springer -Verlag, 1984.

[15b] *Harish-Chandra Collected Papers*, Vol. IV., [1977b] [1977b'], Springer -Verlag, 1984.

[16] *Harish-Chandra Collected Papers*, Vol. III., [1963], Springer -Verlag, 1984.

[17] *Harish-Chandra Collected Papers*, Vol. III., [1968a], Springer -Verlag, 1984.

[18] Harish-Chandra, *My association with Professor Dirac*, in *Reminiscences about a great physicist: Paul Adrien Maurice Dirac*, 34–36, B. N. Kursunoglu, and E. P. Wigner, (Eds.), Cambridge Univ. Press, Cambridge, 1987.

[19] *The Mathematical Legacy of Harish-Chandra*, R. S. Doran, and V. S. Varadarajan, (Eds.), Amer. Math. Soc., Proc. Symp. in Pure Math., Vol. 68, 2000.

[20a] *Harish-Chandra Collected Papers*, Vol. I., [1951f], Springer -Verlag, 1984.
[20b] *Harish-Chandra Collected Papers*, Vol. I., [1952], Springer -Verlag, 1984.
[21] *Harish-Chandra Collected Papers*, Vol. I., [1954e], Springer -Verlag, 1984.
[22] *Harish-Chandra Collected Papers*, Vol. I., [1956a], Springer -Verlag, 1984.
[23] H. Weyl, *Gesammelte Abhandlungen*, Band I, 248–297. Springer-Verlag, 1968.
[24] *Harish-Chandra Collected Papers*, Vol. I., [1958a] [1958b], Springer -Verlag, 1984.
[25] *Harish-Chandra Collected Papers*, Vol. I., [1968b], Springer -Verlag, 1984.
[26a] *Harish-Chandra Collected Papers*, Vol. I., Springer -Verlag, 1984.
[26b] V. S. Varadarajan, *Harmonic Analysis on Real Reductive Groups*, Springer Lecture Notes in Mathematics No., 576, Springer-Verlag, 1977.
[27a] *Lie Theory Harmonic Analysis on Symmetric Spaces – General Plancherel Theorems*, J-P. Anker, B. Orsted (Eds.), Progress in Math. 230, Birkhauser, Boston, 2005.

See in the same volume
[27b] E. P. Van den Ban, *The Plancherel theorem for a reductive symmetric space*, loc. cit., 1–97; see also http://www.math.uu.nl/people/ban/manus/banesgt.ps

CPSIA information can be obtained at www.ICGtesting.com
Printed in the USA
236273LV00005B/77/P

V.S. Varadarajan

Reflections on Quanta, Symmetries, and Supersymmetries

Unitary representation theory has great intrinsic beauty which enters other parts of mathematics at a very deep level. In quantum physics it is the preferred language for describing symmetries and supersymmetries. Two of the greatest figures in its history are Mackey and Harish-Chandra. Their work (to use the words of Weyl) affords shade to large parts of present day mathematics and high energy physics. It is to their memory that this volume is lovingly dedicated.

The essays in this volume are like a stroll through a garden of ideas of this rich subject: quantum algebras, super geometry, unitary supersymmetries, differential equations, nonarchimedean physics, are a few of the topics encountered along the way. The author, whose mathematical education evolved out of his interactions with Mackey and Harish-Chandra, concludes this volume with brief portraits of their work, embedded in the context of personal reminiscences.

Mathematics

▶ springer.com